Flood Hazard Characteristics at the Global Scale:
An Observation-based Perspective

Hong Xuan Do

Master of Engineering (Environmental Management)

Thesis submitted to The University of Adelaide,

Faculty of Engineering, Computer and Mathematical Sciences,

School of Civil, Environmental and Mining Engineering

in fulfilment of the requirements

for the degree of Doctor of Philosophy

Contents

Abstract

Understanding large-scale flood hazard characteristics using streamflow observations is important for designing effective mitigation and adaptation strategies that reduce the future impacts of floods. Barriers to understanding floods include fragmented (and even conflicting) scientific findings of regional studies, limited spatiotemporal coverage of streamflow observations, and the complexity of flood generating processes. This thesis aims to improve the observation-based understanding of flood hazard characteristics at the global scale, focusing on three key objectives: (1) collating streamflow databases to support global-scale hydrological research; (2) identifying global patterns of flooding characteristics using the most comprehensive available streamflow database(s); and (3) evaluating the ability of hydrological simulations to reproduce trends exhibited from streamflow observations.

An important element of this research was the production of the Global Streamflow Indices and Metadata (GSIM) archive. GSIM was initiated to develop an unprecedented daily streamflow dataset containing more than 30,000 stations worldwide by compiling 12 free-to-access databases. Significant efforts were invested in developing GSIM to produce a comprehensive set of metadata (e.g. catchment identifiers, catchment boundary, and landscape attributes), to evaluate the quality of the streamflow records that are included, and to derive time series of indices capturing essential aspects of streamflow regimes. GSIM data products have all been made available through a public data repository to support hydrological research.

The first empirical investigation of flood characteristics at the global scale in this research focused on changes in magnitude of annual maxima streamflow, using more than 6000 daily streamflow of the Global Runoff Data Base, the core database of GSIM. The investigation assessed the significance of trends using the Mann-Kendall test coupled with a bootstrapping field significance approach. Across most experiments, there were more stations with significant decreasing trends than significant

increasing trends, indicating limited evidence to support the hypothesis of increasing trends in flood hazard globally. The detected trends were assessed in the context of upstream catchment attributes and the findings suggested a substantial influence of catchment size on changes in floods.

The data arising from the GSIM project was used for a global-scale assessment of flood seasonality to identify homogeneous regions of flood producing mechanisms. The identified relationships of flood generation were then used to predict flood timing across the globe, including ungauged locations, using climate indices derived from atmospheric reanalysis data-products. GSIM was also used as an input to compare observed trends in streamflow extremes to trends identified from simulated discharge of six global hydrological models (GHMs) across more than 3,000 sites globally. The findings showed moderate capacity of GHMs in reproducing spatial pattern of trends, suggesting the usefulness of GHMs in assessing the widespread pattern of changes in flood hazard.

The release of GSIM enables new opportunities for advances in hydrology research through better spatiotemporal coverage of observations. In addition, the observation-based investigations in this research have yielded important findings and represent a significant contribution to improved understanding of flood characteristics at the global scale.

To my beloved wife, Chuc Thi Dang, and daughter, Minh Dang Hong Do,
for always accepting the tantrum of a grumpy graduate
with special sympathy and unwavering love

Statement of Originality

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

Do, H.X., Westra, S., & Leonard, M. (2017). A global-scale investigation of trends in annual maximum streamflow. Journal of Hydrology, 552, 28-43.

Do, H.X., Gudmundsson, L., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata archive (GSIM)-part 1: the production of a daily streamflow archive and metadata. Earth System Science Data, 10(2), 765-785.

Do, H.X., Westra, S., Leonard, M., and Gudmundsson, L., Global-scale prediction of flood timing using atmospheric reanalysis (submitted to Water Resources Research).

Do, H.X., Zhao, F., Westra, S., Leonard, M., and Gudmundsson, L., Historical and future changes in global flood magnitude – Evidence from a model-observation investigation (for submission 06/2019).

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: Date: 03/05/2019

Acknowledgements

I would like to express my deep appreciation to my supervisors, Prof. Dr. Seth Westra and Dr. Michael Leonard, for their ongoing support during my challenging but rewarding PhD journey. In hindsight, their patience and assistance has gradually transformed a PhD candidate into a well-trained hydrologist with strong connection to the global scientific community. I am particularly in debt to the excellent guidance of Seth, not only in terms of scientific aspects, but also on how to initiate collaboration with other scholars, and to communicate effectively to a broad community. His tough, but insightful, comments have always reminded me that excellent research requires strong commitment, work ethic, discipline and professionalism. I am also grateful to Michael for his generosity in sharing his extensive knowledge and technical experience, which has helped me to overcome many technical barriers throughout my four-year candidature. The development of the Global Streamflow Indices and Metadata (GSIM) archive, one of the key elements of my thesis, has not been possible without hints and suggestions from Michael. It has been a pleasure to have Seth and Michael guiding me through my PhD journey.

I would like to show my appreciation for Dr. Lukas Gudmundsson from ETH Zurich, not only for his bright ideas during the GSIM project, but also for the many precise and constructive comments on our joint publications. I deeply appreciate the insightful discussions and the hospitality of Prof. Dr. Sonia Isabelle Seneviratne (ETH Zurich), Prof. Dr. Albert van Dijk (Australia National University), Prof. Dr. Hayley Fowler (Newcastle University upon Tyne) and Dr. Fang Zhao (Potsdam Institute for Climate Impacts Research) who have hosted me as a visiting researcher at their institutions.

This work would not have existed without the Australia Awards Scholarship, which has provided me with an once-in-a-lifetime opportunity to pursue a PhD. Many of my international travels would not have been possible without the administrative support from Ms. Niranjala Seimon (the

International Student Centre – University of Adelaide), and the traveling grants sponsored by the University of Adelaide.

Thanks are due to Leticia Mooney, Alison-Jane Hunter and Owen Lindsay, the professional editors who provided editing services for the papers constituting this thesis, according to the guidelines laid out in the university-endorsed, national *Guidelines for editing research theses*.

I would like to thank all my friends in Adelaide, for making my PhD journey a joyful one through the time they shared with me. Sincere thanks are also due to all staff members at the School of Civil, Environmental and Mining Engineering, University of Adelaide, for their excellent support to every inquiry of mine.

I am eternally grateful to my family for always being by my side. The unwavering love from my wife, Chuc Thi Dang, and my daughter, Minh Dang Hong Do, has raised me up during difficult times and assured me a place where I truly belong. My parents and sisters in Vietnam, with all of their support throughout my candidature, are the hidden strength behind my every success.

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Background and research objectives

Flooding is one of the leading natural disasters in terms of human fatalities and economic loss at the global scale (CRED, 2015, Guha-Sapir, 2014, Kousky, 2014, Miao, 2018, Smith, 2003). In recent years, rising trends in the population impacted by floods (Munich Re, 2004, Munich Re, 2015) and global economic losses due to floods (Mills, 2005, Kundzewicz et al., 2013) have raised concern of floods becoming more severe in a warming climate. However, the complexity of the processes leading to flood risk, which is determined by not only flood hazard but also flood exposure and flood vulnerability (Kron, 2005), indicates a limitation of using trends in reported flood damage (e.g. number of fatal and economic loss due to floods) to interpret trends in flood hazards (e.g. the extent, magnitude or frequency of peak-flow events). For instance, the rapid growth of population and assets exposed to floods (Hallegatte et al., 2013, Hanson et al., 2011, Hirabayashi et al., 2013, Najibi and Devineni, 2018) indicates that a 'no-trend' scenario in flood hazards might generate an increasing flood risk, assuming no change regarding flood vulnerability. Therefore, in order to understand the drivers of change to floods, it is critical to directly investigate historical changes in the physical manifestation of floods (i.e. the 'hazard') and attribute detectable changes to the physical mechanisms driving floods. An improved understanding of flood hazard characteristics at the global scale will enable evidence-based projections of future floods hazards, which has substantial importance in designing effective mitigation and adaptation strategies to reduce flood risks in a changing climate.

In-situ streamflow observations are an important asset for obtaining insights of global flood hazard characteristics. The numerous analyses of streamflow observations at the continental and global scale have generated important scientific evidence of the key factors driving flood hazards and their influence on trends in floods in a warming climate. In regions where extreme rainfall

plays the dominant role, changes in floods are known to be highly correlated to changes in extreme precipitation (Blöschl et al., 2017, Mediero et al., 2015, Mallakpour and Villarini, 2015), which have experienced an overall increasing trend attributable to anthropogenic climate change (Westra et al., 2013, Guerreiro et al., 2018, Donat et al., 2013, Min et al., 2011). In some mountainous and high-latitude regions, where snowmelt-related processes dominate flood generation, flood magnitude was found to decrease in a warmer climate, potentially due to earlier snowmelt timing (Burn and Whitfield, 2016, Cunderlik and Ouarda, 2009) or a shift from snowfall to rainfall (Ye et al., 2017, Berghuijs et al., 2014). However, over regions with a significant relationship to antecedent soil moisture, the implication of climate change on flood hazard remains unclear due to the combined effect of changes in potential evaporation and extreme precipitation (Bennett et al., 2018, Wasko and Sharma, 2017, Ivancic and Shaw, 2015).

Empirical (data-driven) studies analysing whether large streamflow events are changing over time have been conducted at the global (Kundzewicz et al., 2004, Milly et al., 2002, Hodgkins et al., 2017), continental (Hall et al., 2014, Mangini et al., 2018, Bormann et al., 2011) and regional (Delgado et al., 2010, Zhao et al., 2015, Zhang et al., 2001, Petrow and Merz, 2009) scales. These studies have generally detected a mixed pattern of trends, with some global-scale studies finding more stations showing decreasing trends than increasing trends (Hodgkins et al., 2017, Kundzewicz et al., 2004); an unexpected result given the historical increases in flood risk as well as the documented increases in heavy rainfall events described earlier. Furthermore, over some flood-prone regions such as South and East Asia, observational studies have not yet generated a coherent picture about changes in floods and the underlying mechanisms due to limited data coverage (Hannah et al., 2011, Gupta et al., 2014). It is therefore difficult to derive a global perspective on flood hazard characteristics using observation-based investigations alone.

An alternative approach to obtaining insights of flood hazard characteristics at the global scale is through simulations of large-scale hydrological models (Arnell and Gosling, 2014, Dankers et al., 2014). The key benefit of this approach is the global coverage of streamflow simulations with the possibility

to extend temporal coverage back to the pre-industrial period or forward to the end of the 21st century. With these important features, global hydrological models (GHMs) have been increasingly used to assess the implications on flood hazards of climate change and other anthropogenic factors such as land use change or river morphology. For instance, GHMs have been used to assess climate change impacts on fluvial flooding magnitude (Dankers et al., 2014) or frequency (Arnell and Gosling, 2014) at a 0.5 × 0.5 arc-degree spatial resolution. However, there are several challenges when using GHMs to understand flood hazard characteristics, including the tremendous number of required inputs representing climate and land-surface characteristics (Chang and Franczyk, 2008, Winsemius et al., 2013, Ward et al., 2015); acknowledged high uncertainty in streamflow simulations due to uncertainties of model inputs (Bierkens, 2015, Ward et al., 2013, Wood et al., 2011), model structure (Gupta et al., 2014, Schaake et al., 2006), or parameterisation of human influence (Bierkens et al., 2015, Veldkamp et al., 2018). The implications of these challenges are that simulated trends in flood hazards could be biased due to model uncertainty (Giuntoli et al., 2018).

Another challenge of using GHMs to understand flood hazard characteristics is the contradictory performance of GHMs in terms of reproducing trends in streamflow indices relative to observation-based studies. In particular, investigations of global flood hazard using GHMs simulations (Dankers et al., 2014, Arnell and Lloyd-Hughes, 2014, Arnell and Gosling, 2014, Alfieri et al., 2015, Alfieri et al., 2017) found more increasing trends occurred in land grid cells than decreasing trends, which were attributed to changes in climate variables simulated by general circulation models (Bouwer, 2013, DeGaetano and Castellano, 2017, Kharin et al., 2007). However, the reliability of these conclusions is difficult to determine, owing to the fact that no evaluation study has been conducted to assess model capacity in reproducing trends.

As a result of limitations to both empirical and model-based studies, it is still not possible to draw a common conclusion about changes to flood hazard characteristics at the global scale. The findings of both model-based and observation-based studies are still fragmented, and even conflicting, due to the

varying strategies in choosing study samples, domain, analysis methodology, and reference period. This thesis, therefore, aims to contribute an observation-based perspective to further improve understanding of flood hazard characteristics at the global scale. The present thesis takes advantage of the most comprehensive set of streamflow observations together with the latest global datasets to achieve three objectives:

Objective 1: To compile a global streamflow database with improved spatiotemporal coverage using free-to-access data sources. The dataset, which has been made freely accessible in the public domain, can be used to support advances in large-scale hydrological research.

Objective 2: To complement the current understanding of trends in flood magnitude and the underlying mechanism.

Objective 2.1: To obtain new insights of trends in flood hazard using the most comprehensive streamflow databases. These findings are important to complement the global picture of changes in flood hazard.

Objective 2.2: To identify the key natural and anthropogenic factors driving flood occurrence across the globe. This understanding is of scientific importance to explain the underlying mechanisms of detectable trends in the Anthropocene.

Objective 3: To evaluate the capacity of global hydrological models in terms of reproducing trends in flood hazards. The evaluation is based on a large-sample of streamflow observations to obtain insights about the credibility of using GHMs to assess climate change implications on trends in flood hazards.

1.2 Thesis overview

This thesis is organized into six chapters, where the key publications constitute Chapters 2 to 5. The other publications are included as supplementary materials: they have been contributed to substantially by the thesis author and are also an important element of this research as they directly contribute to achieving the defined objectives. Details about the publications within this thesis are outlined in Table 1-1 while Figure 1-1 illustrates the synthesis of the

main-chapters and the supplementary materials, each is a paper/dataset that has been either published or submitted to peer-reviewed journals.

Table 1-1: Detail of publications within this thesis.

Item	Title	Contribution	Publication status
Paper 1 (Chapter 2*)	The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1	Main author 85%	https://doi.org/10.5194/essd-10-765-2018
Paper 2 (Chapter 3*)	A global-scale investigation of trends in annual maximum streamflow	Main author 85%	https://doi.org/10.1016/j.jhydrol.2017.06.015
Paper 3 (Chapter 4*)	Global scale prediction of flood timing using atmospheric reanalysis	Main author 85%	Under review (Water Resources Research)
Paper 4 (Chapter 5)	Historical and future changes in global flood hazard – Evidence from a model-observation investigation	Main author 80%	For submission 06/2019
Supplementary Dataset 1	The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1	Main author 90%	https://doi.pangaea.de/10.1594/PANGAEA.887477
Supplementary Dataset 2	The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2	Co-author 25%	https://doi.pangaea.de/10.1594/PANGAEA.887470
Supplementary Article 1	The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2	Co-author 30%	https://doi.org/10.5194/essd-10-787-2018
Supplementary Article 2	Observed Trends in Global Indicators of Mean and Extreme Streamflow	Co-author 10%	https://doi.org/10.1029/2018GL079725
Supplementary Article 3	Large-sample hydrology: recent progress, guidelines for new datasets and outstanding challenges	Co-author 20%	Under review (Hydrological Sciences Journal)

* This paper has been reformatted and renumbered as a Chapter within this thesis, but the content is identical to the published (Appendix A and Appendix B) or submitted versions.

The remainder of the thesis is organised as follows:

Chapter 2 addresses the limited data coverage of the Global Runoff Data Base (GRDB; Objective 1), arguably the most popular daily streamflow database for global hydrological studies. Limited spatiotemporal coverage of the GRDB is the key barrier to generating meaningful understanding about flood characteristics for several regions such as the South and East Asia. The Global Streamflow Indices and Metadata (GSIM) archive, using the GRDB as the core database, is compiled from daily streamflow records of 12 freely available databases, yielding an unprecedented number of streamflow gauges, representing more than 30,000 locations worldwide. Significant efforts were invested in developing GSIM to produce comprehensive metadata (e.g. catchment identifiers, catchment boundary, and landscape attributes such as climate type, land cover type, the human population). The GSIM catalogue (**Supplementary Dataset 1:** contains stream-gauge metadata, catchment boundary and extracted land-scape attributes) has been made available through PANGAEA, a public data repository, to support large-scale hydrological research.

Chapter 3 presents a global-scale investigation of changes in the magnitude of annual maxima streamflow (Objective 2.1) using more than 6000 daily streamflow gauges of the GRDB (which was undertaken prior to development of the GSIM archive). Across most experiments, there were more stations with significant decreasing trends than significant increasing trends, indicating limited evidence to support the hypothesis of increasing trends in flood hazard globally. The detected trends were assessed in the context of upstream catchment attributes (e.g. drainage area, changes in forest cover, the presence of dams and reservoirs) and the findings suggested a substantial influence of catchment size on changes in floods (Objective 2.2). The key assessment of this paper was updated using GSIM, and the results presented in Chapter 5 yielded similar conclusions regarding changes in flood hazard at the global scale.

Chapter 4 presents a comprehensive analysis of flood seasonality at the global scale using the GSIM archive as the observational basis. The spatial association of flood timing is compared against seven climate predictors to identify the spatial distribution of the generation mechanism of streamflow maximums at the global and continental scales (Objective 2.2). The identified relationship of flood generation is then generalised across both gauged and ungauged locations to enable a global map of flood timing.

Chapter 5 provides an evaluation of global hydrological models (available through the Inter-Sectoral Impact Model Intercomparison Project) in terms of reproducing trends observed across more than 3,600 sites globally (Objective 3). The findings show the moderate capacity of GHM simulations forced with observational climate data in reproducing the key characteristics of trends (e.g. mean, spread, field significance and spatial pattern). The performance suggests the usefulness of GHMs in assessing the widespread pattern of changes in flood hazard. Streamflow simulations forced with hindcast climate data obtained from general circulation models, however, have significantly lower performance, indicating the influence of climate model uncertainty on GHMs in terms of reproducing trends (Objective 3). Projections of discharge under two socio-economic scenarios are also assessed, highlighting the substantial influence of a warming climate to changes in flood hazard (Objective 2.2).

Chapter 6 concludes this thesis with a summary of the key research findings, each contributing directly to a defined objective of this thesis. Limitations that have not been fully addressed are also presented in this concluding chapter, together with future directions to further improve the understanding of flood hazard at the global scale.

The Supplementary section comprises two data sets together with three additional publications arising from this thesis, which also contribute substantially to the defined objectives. **Supplementary Article 1** (published on Earth System Sciences Data) describes the evaluation for streamflow quality of all GSIM records and the production of time series of indices capturing essential aspects of streamflow regimes (Objective 1). The indices described in Supplementary Article 1 are also made available through the PANGAEA data repository (**Supplementary Dataset 2**) and can be freely redistributed without any restriction. **Supplementary Article 2** (published in Geophysical Research Letters) takes advantage of the freely available indices to assess changes in low-/mean-/high-flow indicators at the global scale (Objective 2.1). The results highlight increasing trends in flood indicators over some regions, indicating a shift of the whole flow distribution to a wetter condition. **Supplementary Article 3** (for submission 03/2019 to the Hydrological Sciences Journal Special Issue on "Hydrological Data: Opportunities and Barriers") focuses on recent progress and challenges of producing and exchanging large-sample hydrological datasets together with guidelines toward better data availability for advances in large-sample hydrological investigations (Objective 1).

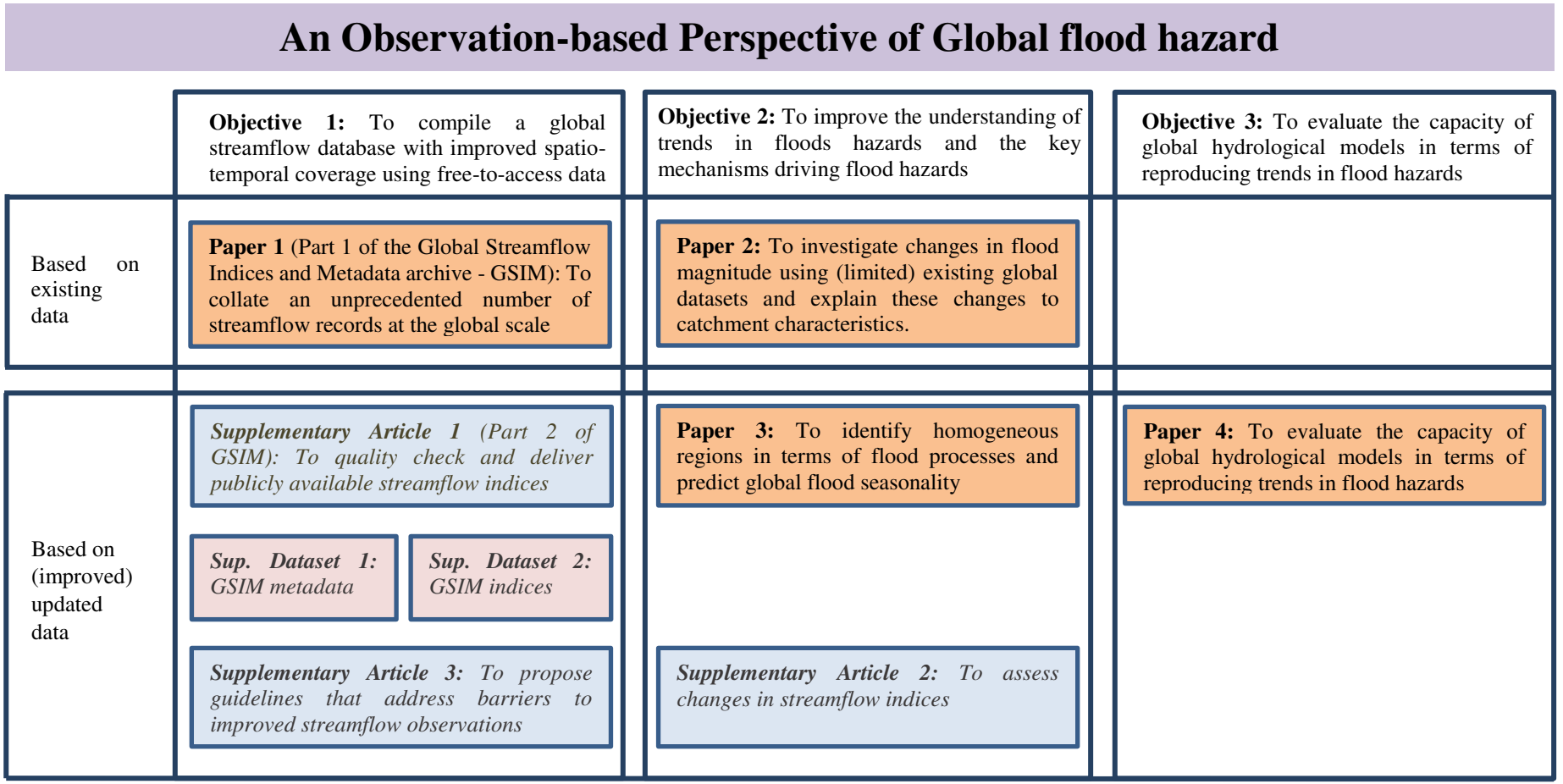


Figure 1-1: The link of each paper to thesis objectives. Paper 2 was finalised prior to the development of GSIM (key assessment was reproduced using GSIM in Chapter 5).

Chapter 2. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of daily streamflow archive and metadata (Paper 1)

Statement of Authorship

Title of Paper	The Global Streamflow Indices and Metadata archive (GSIM) - Part 1: the production of a daily streamflow archive and metadata.
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Do, H.X., Gudmundsson, L., Leonard, M., Westra, S. (2018). The Global Streamflow Indices and Metadata archive (GSIM) - Part 1: The production of a daily streamflow archive and metadata. Earth System Science Data, 10(2), 765-785.

Principal Author

Name of Principal Author (Candidate)	Hong Xuan Do
Contribution to the Paper	Designed the framework of the paper, collected and compiled data, produced data products and descriptors to be made available online, wrote manuscript and acted as corresponding author.
Overall percentage (%)	85
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 08/02/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Lukas Gudmundsson
Contribution to the Paper	Suggest scope of the paper, helped to develop Rscript for data compiling, provided feedbacks on manuscripts and responses to reviewers.
Signature	Date 11/02/2019

Name of Co-Author	Michael Leonard
Contribution to the Paper	Suggest scope of the paper, provided feedbacks on manuscripts and responses to reviewers.
Signature	Date 19/02/2019

Name of Co-Author	Seth Westra

Contribution to the Paper	Suggest scope of the paper, provided feedbacks on manuscripts and responses to reviewers.

Signature		Date	15/02/19

Abstract.

This is the first part of a two-paper series presenting the Global Streamflow Indices and Metadata archive (GSIM), a worldwide collection of metadata and indices derived from more than 35,000 daily streamflow timeseries. This paper focuses on the compilation of the daily streamflow timeseries based on 12 free-to-access streamflow databases (seven national databases and five international collections). It also describes the development of three metadata products (freely available at https://doi.pangaea.de/10.1594/PANGAEA.887477): (1) a GSIM catalogue collating basic metadata associated with each timeseries, (2) catchment boundaries for the contributing area of each gauge, and (3) catchment metadata extracted from 12 gridded global data products representing essential properties such as land cover type, soil type, climate and topographic characteristics. The quality of the delineated catchment boundary is also made available and should be consulted in GSIM application. The second paper in the series then explores production and analysis of streamflow indices. Having collated an unprecedented number of stations and associated metadata, GSIM can be used to advance large-scale hydrological research and improve understanding of the global water cycle.

2.1 Introduction

Streamflow observations with global coverage are essential to progress the science of large-scale hydrology. For example, global datasets provide particularly value when evaluating global hydrological models (Gudmundsson et al., 2012a, Ward et al., 2013, Huang et al., 2016), producing runoff estimation data-products (Fekete et al., 2002a, Fekete et al., 2002b, Vörösmarty et al., 1989, Gudmundsson and Seneviratne, 2016), investigating large-scale weather patterns and their relation to hydrological extremes (Ward et al., 2014, Wanders and Wada, 2015), and detecting changes in the global hydrological over space and time (Milly et al., 2002, Kundzewicz et al., 2012, Do et al., 2017, Gudmundsson et al., 2017), amongst numerous other applications.

Despite the fundamental, widespread and varied applications that streamflow observations support, there are many obstacles to the existence and utility of a large-scale streamflow archive. Firstly, there are threats to the quantity of data, such as political sensitivities (Nelson, 2009), cost recovery and strict access policies (Hannah et al., 2011), unavailability in an electronic format, consistency of data formats, limited documentation, missing metadata, and a lack of resources for database maintenance and updating. Secondly, there are difficulties associated with the quality of the data in many regions, such as poor spatial coverage, poor quality control, variable quality control between regions, inconsistent metadata, imprecise geographic coordinates of the site, changes in the density of stream gauges and variable record lengths. Lastly, even in locations where there are abundant and high quality streamflow observations, there can be questions over its utility in specific research such as climate sensitivity analysis due to the manifestation of human impacts—for example, urbanization, land-use changes, channelization and upstream dams (Hannah et al., 2011).

To date, the Global Runoff Data Base (GRDB) maintained by the Global Runoff Data Centre has been the primary dataset used in large-scale hydrological studies, with more than 9000 stations available to the research community (GRDC, 2015). The Global Runoff Data Centre (GRDC) database operates under the auspices of the UN – World Meteorological Organization (WMO), and its database is supported on a voluntary basis so that the number of data

submissions depends on contributions by national authorities. However, although numerous countries have databases of acceptable quality, data supply remains resource intensive and the GRDB remains sparse in some regions. For example, the latest catalogue of the GRDB database (version December 5th 2017) shows that out of 7,238 daily timeseries, there are only 637 stations over South America and only 642 stations over Asia. Moreover, many stations in regions such as Asia and Russia have not been updated for many years and are missing otherwise available data at the end of their records.

The Global Streamflow Indices and Metadata (GSIM) project has been initiated in order to address the demand for a global streamflow database (Bierkens, 2015, Hannah et al., 2011, Merz et al., 2012a, Kundzewicz et al., 2013, Milly et al., 2015, Fekete et al., 2015). The approach of this project is not to collect high-quality data from referenced hydrological networks, which have been conducted in other studies (Hannaford and Marsh, 2006, Whitfield et al., 2012, Burn et al., 2012, Hodgkins et al., 2017, Addor et al., 2017) to support research that requires assumptions regarding the minimum impact of human interference to streamflow, such as the investigation of climate change implication on changes in extreme events. Instead, the activities of the GSIM project have been to collate publicly available data, apply basic consistency to the formatting and establish a standardised set of metadata. In so doing, GSIM intends to promote more widespread use of streamflow data, facilitate improved research outcomes through increased spatial coverage and gauge density, and tackle ongoing challenges for the hydrological community; for example, addressing fundamental issues of data quality, identifying additional data sources, lobbying for continuity of data networks and developing a method for improved governance and maintenance of streamflow data at the global scale.

To maximise the value of the streamflow dataset for a wide range of applications, the GSIM project also seeks to provide information on catchment characteristics upstream of the streamflow gauging station. This necessitates a consistent approach to delineating the upstream catchment boundary for every gauge station, and this is achieved using data from a global Digital Elevation Model (DEM). This is because, with the exception of the GRDB databases, catchment boundaries representing the direct drainage area of stations were

unavailable. Filling in this missing element of metadata is important to facilitate further analysis of the streamflow observations with respect to a wide and ever-increasing variety of spatial datasets. Although there have been previous efforts in producing catchment boundaries for a smaller number of stations (Lehner, 2012, Arsenault et al., 2016, Schaake et al., 2006, Addor et al., 2017), similar work at this magnitude has not been undertaken. This task is complicated by a lack of precision in the supplied geographic coordinates of a given site; for example when a catchment boundary is extracted, the corresponding calculated area may not match the reported area of the catchment and a procedure for checking minor shifts in the coordinates is needed to improve identification of the likely catchment boundary. The quality of the delineated catchment boundary is also made available to GSIM users and should be considered prior to using this data product and any accompanied information.

The availability of catchment boundaries for each gauge enables the association of environmental variables to each gauge by extracting them from corresponding global-scale gridded products. As part of the GSIM project, a number of global data products are provided as an additional dataset so that a user can readily filter the GSIM dataset according to specific interests; for example, by climate type, soil type, land-use type, irrigation area and population density. Other potential applications of this auxiliary information might include a comparison to a database of dams for identifying upstream impacts; to remotely sensed estimates of forest cover or urban extent for determining land-use change; to population demographics for improving estimates of flood exposure; and to hydrological model outputs for evaluating model performance.

Finally, to facilitate benefits of this project to the broader community, indices characterising water-balance aspects, hydrological extremes and features of the seasonal cycle have been derived from the GSIM timeseries and will be made publicly available. To ensure standardised quality for the derived indices, a quality control procedure coupling the information provided by data providers and a data-driven approach was also applied.

This is the first paper of a two-part series detailing the production of GSIM and corresponding data products. This paper outlines the provenance of daily streamflow timeseries (Section 2.2), procedures for reformatting and combining

the timeseries (Section 2.3), the development of metadata associated with each gauge (Section 2.4), an overall summary of the GSIM timeseries and metadata (Section 2.5), and data availability (Section 2.6). As the time series database cannot be made available online due to varieties of terms and conditions from data providers, the second paper in this series (Gudmundsson et al., 2018b) is dedicated to the production of streamflow time series indices, including: (1) checks for data quality, (2) the production of streamflow timeseries indices, and (3) homogeneity assessment of the derived indices.

2.2 Daily streamflow data and where to find them

GSIM is a compilation of 12 databases that have either open-access or restricted-access policies, and that collectively represent a total of 35,002 stations. The spatial distribution and the number of stations available in each database is illustrated in Figure 2-1 (continental scale figures are also provided as supplementary materials). A summary of the data sources is also provided in Table 2-1 and detailed information on each database is elaborated upon in following sections. The list of databases identified as part of GSIM is not exhaustive of all possible data sources, only of those that were known to the authors and readily accessible within the project time frame. Where additional data is available in a convenient format it may be possible to further augment GSIM in the future.

The various data sources were classified as either a “research database” or a “national database”. The reasons for this classification are further outlined in Section 2.3, but relate to issues when merging databases and removing duplicate gauges. The data sources include:

(1) **Research databases:** Databases with daily streamflow data that have been compiled on an ad hoc basis from a variety of original sources by research organisations. This category includes five different databases: The Global Runoff Data Base (GRDB); the European Water Archive (EWA); the China Hydrological Data Project (CHDP) data archive; the GEWEX Asian Monsoon Experiment – Tropics (GAME) data archive; and the Regional Hydrographic Data Network for the Arctic Region (ARCTICNET) data archive.

(2) **National databases:** Databases with daily streamflow data made publicly available by national water authorities as part of water-related regulations. This category includes seven databases: The USGS Water Data for USA database (USGS); Canada's National water data archive (HYDAT); Japan's Ministry of Land and Infrastructure database Water Information System (MLIT); Spain's digital hydrological year book database (Anuario de aforos digital 2010–2011, AFD); Australia's Bureau of Meteorological Water Data Online database (BOM); India's Water Resources Information System database (WRIS); and Brazil's National Water Agency database (ANA).

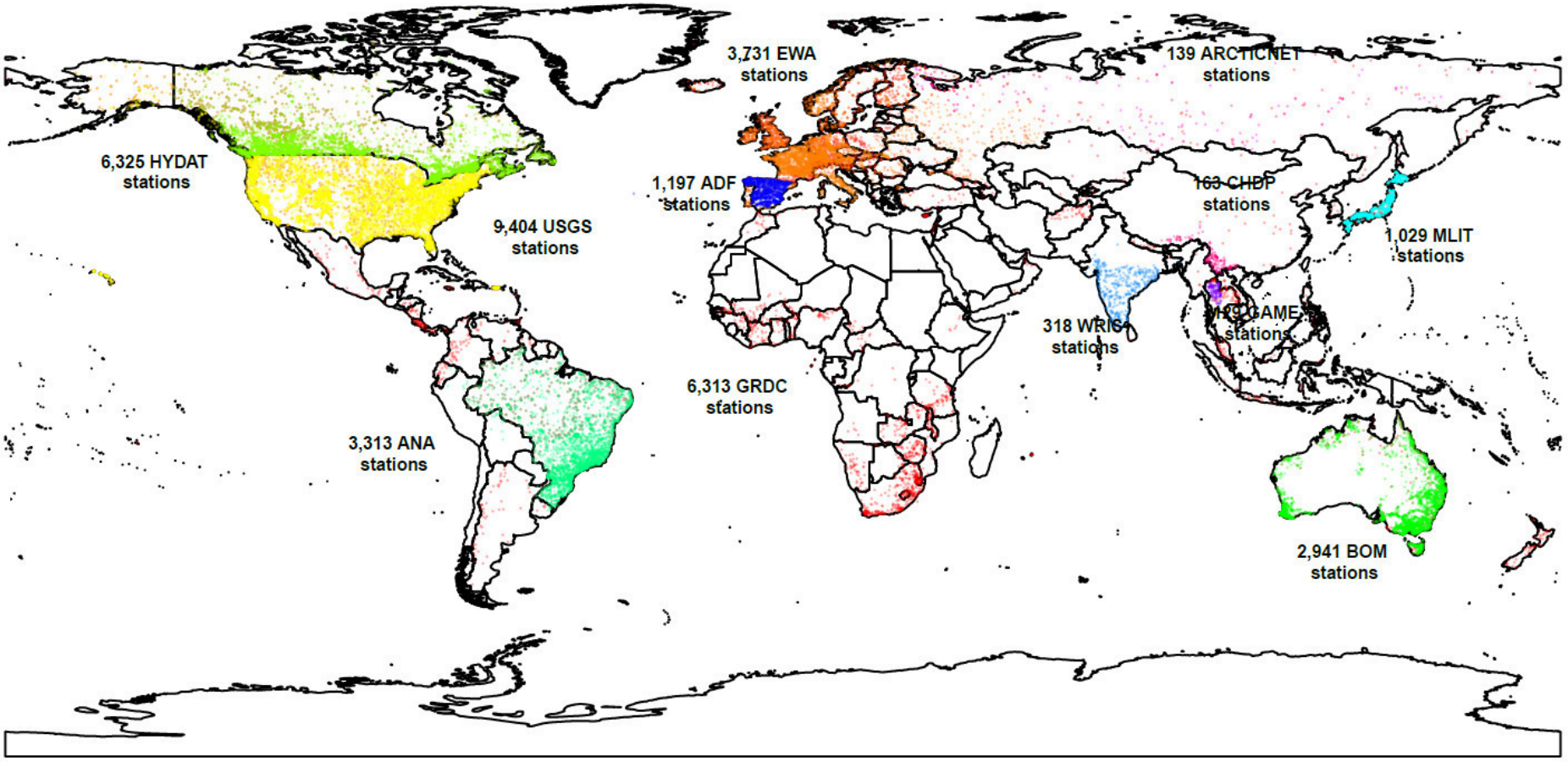


Figure 2-1: The distribution of stations from original data sources.

Table 2-1: Basic information of daily streamflow databases included in GSIM project

Database (referred name)	Database category	Spatial coverage	Data access information
Global Runoff Data Base (GRDB)	Research database	Global	www.bafg.de/GRDC/ Archived database can be obtained via written request to the Global Runoff Data Centre. This database is updated when new data submitted by national suppliers.
European Flow Regimes from International Experimental and Network Data (EWA)	Research database	European	http://ne-friend.bafg.de/servlet/is/7413/ Data can be obtained via written request to the Global Runoff Data Centre . This database has been frozen since October 2014 and is being integrated into GRDB database.
A Regional, Electronic, Hydrographic Data Network for Russia (ARCTICNET)	Research database	Russia	http://www.russia-arcticnet.sr.unh.edu/ Archived and closed historic database. Part of this data-archived have been included in the databases of the Global Runoff Data Centre and updated based on data deliveries
China Hydrology Database Project (CHDP)	Research database	China	http://www.oberlin.edu/faculty/aschmidt Archived and closed historic database can be obtained via written request to the author of the database
GEOSS ana MAHASRI Experiment in Tropics (GAME)	Research database	Thailand	http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/routine/rid-river/disc_d.html Archived and closed historic database
U.S. National Water Information System (USGS)	National database	USA	http://waterdata.usgs.gov/nwis Individual timeseries can be downloaded from the data portal (updated regularly)
Canada National Water Data Archive (HYDAT)	National database	Canada	https://ec.gc.ca/rhc-wsc/ Archived database. The archive is updated quarterly by the data authority.
Brazil National Water Agency (ANA)	National database	Brazil	http://hidroweb.ana.gov.br/ Individual timeseries can be downloaded from the data portal (updated regularly)
Japan Water Information System (MLIT)	National database	Japan	http://www1.river.go.jp/ Individual timeseries can be downloaded from the data portal (updated regularly)
Anuario de aforos digital 2010 - 2011 (AFD)	National database	Spain	http://ceh-flumen64.cedex.es/anuarioaforos Archived database, DVD available from Spanish authorities (updated annually)
Australia Water Data Online (BOM)	National database	Australia	http://www.bom.gov.au/waterdata/ Individual timeseries can be downloaded from the data portal (updated regularly)
Water Resources Information System of India (I-WRIS)	National database	India	http://www.india-wris.nrsc.gov.in/wris.html Individual timeseries can be downloaded from the data portal (updated regularly)

2.2.1 The Global Runoff Data Base (GRDB)

The daily streamflow dataset received from the GRDC (6,313 stations with greater than 10 years record; see also Gudmundsson and Seneviratne (2016)) is referred to as the GRDB in this project. To date, the GRDB has been the largest and most extensively used dataset for streamflow analysis at regional and global scales. It was thus considered as the starting point and "base" for the GSIM project. Indeed, it was awareness of data not available from the GRDB that prompted the initial search for additional sources of data to complement the database.

The GRDC was initiated in 1988 by the WMO and is now maintained at the German Federal Institute of Hydrology in Koblenz. The GRDC provides free and unrestricted access to all hydrological data and products, although the data policy indicates that requests for data must reach the GRDC in written form to ensure data users do not redistribute the timeseries. More detail about the GRDC data policy, and procedure for obtaining its timeseries, are outlined at http://www.bafg.de/GRDC/EN/01_GRDC/12_plcy/data_policy_node.html.

2.2.2 The European Water Archive

The European Water Archive (referred to as the EWA in this paper) is one of the most comprehensive streamflow timeseries archives in Europe, with more than 3000 river gauging stations distributed across 29 countries. This archive is also currently held by the GRDC and available under the GRDC data policy (http://www.bafg.de/GRDC/EN/04_spcldtbss/42_EWA/ewa_node.html, last accessed 03 Jan 2018). The EWA stations used in this paper were selected using the same criteria as Gudmundsson and Seneviratne (2016), with a total of 3,731 daily records.

2.2.3 The China Hydrology Data project

The China Hydrology Data Project (CHDP) aims to digitise an arrangement of hydrological measurements taken at Chinese stations. These measurements (including daily discharge) were originally only available in book form (Henck et al., 2010). The original data were collected by the Chinese Hydrology Bureau and published in annual yearbooks. At the time GSIM began, discharge data were only available for the Yunnan-Tibet International Rivers, which

corresponded to 163 stations until 1987. This project has been terminated since 2000s and thus no further update available. The data and metadata were obtained directly from the author of the project. Detailed information can be viewed at http://www.oberlin.edu/faculty/aschmidt/chdp/index.html (last accessed 23 June 2017).

2.2.4 The GEWEX Asian Monsoon Experiment – Tropics project

The GEWEX Asian Monsoon Experiment – Tropics project (GAME) was initiated in 1996 to monitor several hydro-climatological variables over the humid temperate area in south-east Asia. As one of several important activities in this project, many hydrological observation datasets were collected, including streamflow data. Available streamflow data was provided by the Royal Irrigation Department of Thailand, and comprised 129 timeseries spanning a period from 1980 to 2000. Daily discharge data and associated metadata were archived and can be accessed online at http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/routine/rid-river/index.html (last accessed 23 June 2017).

2.2.5 The ARCTICNET project

A regional hydrometeorological data network for the pan-Arctic Region project is a regional data bank that can be accessed via the internet and is referred to as ARCTICNET in this paper. The databank is designed to support hydrological sciences and water resource assessments over this region with the goal of estimating the contemporary water and constituent balances for the pan-Arctic drainage system. ARCTICNET is a static dataset and some time series have been included in the databases of the GRDC and updated based on data deliveries. Although most data provided in the data portal are at monthly resolution, there are 139 high-quality daily streamflow timeseries across Russia that are also available, which have not been fully integrated into GRDB. Although ARCTICNET's future status is likely to be a part of the GRDB, these stations still have been considered in GSIM production and are referred to as the ARCTICNET database in this paper. These timeseries, along with their metadata, were archived and can be downloaded at http://www.r-arcticnet.sr.unh.edu/v4.0/index.html (last accessed 23 June 2017).

2.2.6 The USGS database

The USGS National Data Services for the US provide access to water resources data collected at approximately 1.5 million sites in all 50 states of the USA, also including the District of Columbia, Puerto Rico, the Virgin Islands, Guam, American Samoa and the Commonwealth of the Northern Mariana Islands. All timeseries and associated metadata can be queried from the data portal http://waterdata.usgs.gov/nwis (last accessed 23 June 2017). To ensure the queried data have sufficient geographic metadata (critical for the present project), the stations listed in the "Geospatial Attributes of Gages for Evaluating Streamflow, version II" (GAGES II) database were used (Falcone, 2011). The timeseries from 9,404 stream gauges obtained from the USGS data portal are referred to as the USGS database in this paper.

2.2.7 The HYDAT database

Canada's National Water Data Archive (HYDAT) is a database containing daily observed hydrometric data from publicly funded gauges in Canada. Also available in the HYDAT database are metadata about the hydrometric stations, such as latitude and longitude, catchment area, record length, as well as information regarding flow conditions (current status, regulated or natural regime). The database is updated four times per year and currently contains data for 6,325 streamflow stations across Canada. The raw data, as well as an extractor executable, are publicly available from Environment Canada's website at https://ec.gc.ca/rhc-wsc/default.asp?lang=En&n=9018B5EC-1 (last accessed 23 June 2017).

2.2.8 The ANA database

The data portal HIDROWEB was organised by the Brazilian National Water Agency (ANA). It provides a database with all the information collected by Brazil's hydrometeorological network. Streamflow data and associated metadata were made publicly available by Brazil's national water regulations, and have been used extensively to monitor critical events, such as floods and droughts. Individual timeseries and their associated metadata can be viewed or downloaded at http://hidroweb.ana.gov.br (last accessed 23 June 2017). The

3,313 stations downloaded from this website are referred to as the ANA in this paper.

2.2.9 The AFD database

Spanish streamflow data were retrieved from the digital hydrological year book (Anuario de aforos digital 2010–2011, AFD), which provides observations until 2013–2014 and is freely accessible online (http://ceh-flumen64.cedex.es/anuarioaforos/default.asp (last accessed 23 June 2017). For the GSIM, we used the timeseries that was used to develop the E-RUN dataset (Gudmundsson and Seneviratne, 2016). The original DVD containing the full database was obtained directly from the Spanish authorities via a written form request. This collection contains streamflow data from 1,197 gauging stations, and is referred to as ADF in this paper.

2.2.10 The MLIT database

In Japan, the Ministry of Land, Infrastructure, Transport and Tourism is responsible for organising hydrological data. All records are disseminated at http://www1.river.go.jp/ (last accessed 23 June 2017). As at 2010, the database kept records of all river stations (at both discharge and gauge level). The composition of the 15-digit station IDs is outlined in the file http://www1.river.go.jp/kitei_sosoku.pdf (PDF), and can be used to query and download timeseries, along with its metadata. As the whole database is recorded in Japanese, the package 'translateR' (Lucas and Tingley, 2016) was used to translate the metadata into English. The timeseries downloaded from the Japanese water data portal (1,029 stations in total) is referred to as MLIT in this paper.

2.2.11 The BOM database

As part of the water reform program established in Australia, Water Data Online was created to provide free access to nationally consistent, current and historical water information. It can be accessed at http://www.bom.gov.au/waterdata (last accessed 23 June 2017). Water Data Online also contains historical data from some stations that are no longer operational. Users can view or download individual streamflow timeseries from the data portal, along with standardised data and reports. The timeseries

measured at 2,941 stations obtained from Water Data Online is referred to as the BOM database in this project.

2.2.12 The WRIS database

The "Generation of Database and Implementation of Web Enabled Water Resources Information System in the Country" project (India-WRIS WebGIS) was initiated as a joint venture of the Indian Central Water Commission (CWC) and the Indian Space Research Organization (ISRO). Unclassified data can be accessed online and free of charge at: http://www.india-wris.nrsc.gov.in/wris.html (last accessed 23 June 2017), while the metadata is documented at: http://www.cwc.nic.in/main/downloads/Hydrological%20network%20details%20of%20CWC.pdf. All 318 stations were downloaded from the website. They are referred to as the WRIS database in this paper.

The production of timeseries and metadata for GSIM comprises several stages due to the range of data formats and significant variation in the quality of metadata across data sources. To ensure GSIM is presented in a transparent manner, the following sections outline procedures that are used to collate the timeseries across (Section 2.3), and to produce the metadata (Section 2.4).

2.3 Procedure for combining databases

Several of the identified data sources share common spatial domains, where typically the research databases may contain a subset of gauges from the national databases. It is therefore important to correctly identify duplicate timeseries when merging the databases. To maximise the quality of combined timeseries and minimise the requirement to combine timeseries, this task is conducted following three sequential steps: Step 1 – pre-processing the data to a common structure; Step 2 – replacing all GRDB stations in countries that have a national database; and Step 3 – identifying remaining duplicates. From the 35,002 gauges, 3,197 (2,958 and 239 gauges from GRDB and EWA databases respectively) were replaced by national databases in Step 2, and 846 cases of 'very likely identical' stations were identified and removed in Step 3, leaving 30,959 'duplication-free' timeseries available in the GSIM.

2.3.1 Pre-processing the timeseries into a singular data structure

One of the major challenges in producing consistent streamflow indices is that data from different sources have different structures and storage formats. For example, the MLIT database divides streamflow records at one location into separate text files, and each file contains streamflow measurements for one year. In comparison, the HYDAT archive includes streamflow measurements from all available stations in a single matrix.

To address the varying standards of data management, the first step in combining the databases was to reformat all the streamflow records to ensure that each timeseries is kept in a consistent format. Using the GRDB as a guide, it was decided to store all data for a given site in a single text file with three columns: a) date of measurement, b) value of measurement and c) original quality flags (if available), and with basic metadata (e.g. station name, ID, etc.) stored in the header of the file. All additionally derived metadata (i.e. from global gridded products) is stored in the station catalogue. The streamflow measurements were also converted into consistent units (cubic meters per second).

Metadata that have special characters in foreign language sources were also pre-processed into the ASCII encoding system. For river names and station names that are recorded in Spanish (ADF) or Portuguese (ANA), the special characters were replaced by plain alphabetic characters using the core function iconv() of the R programming language. For river names and station names that are recorded Japanese characters (MLIT), The R package 'translateR' (Lucas and Tingley, 2016) was used with the Google Translate API for this task. Although there are some limitations related to this toolset (e.g. some Japanese characters remaining untranslated and requiring manual translation; inconsistency in the translated results using the same original Japanese characters), this option was chosen to enable an automated and expedient translation. As a result, any text-related metadata associated with Japanese stations should be treated with care.

2.3.2 Replace the GRDB stations with national databases, if applicable

The streamflow records hosted by the GRDC (the GRDB and EWA databases) are themselves originally provided by national water agencies, and

have been undergone quality control procedures by the GRDC. In cases that the supplied data contain errors, the GRDC informs data suppliers to improve the quality of their database. In term of data availability, time series downloaded directly from the national data portal usually represents the latest version of streamflow observation, and thus it seemed appropriate to replace stations hosted by the GRDC for countries where an equivalent national database was available. While this approach is efficient, there is a potential downside of removing GRDB stations that were not otherwise present in the national data depositories, perhaps due to differences in maintenance of the databases. Nonetheless, the number of stations available in the GRDB and EWA databases is much lower than that available in national databases for all countries (see Table 2-2). As a result of this step, 2,958 stations located in seven countries (Australia, Brazil, Canada, India, Japan, Spain, and the United States) were removed from the GRDB collection. In addition, 239 stations located in Spain were removed from the EWA archive.

Table 2-2: Number of stations in countries where national databases are available

Country	Database		
	EWA	**GRDB**	**National**
Australia	-	358	2941 (BOM)
Brazil	-	439	3313 (ANA)
Canada	-	1029	6325 (HYDAT)
India	-	0	318 (WRIS)
Japan	-	151	1029 (MLIT)
Spain	239	0	1197 (ADF)
United State	-	981	9404 (USGS)

2.3.3 Identify and remove duplicates in research databases

The method of de-duplicating timeseries involves identification of duplicates where two data sources have overlapping coverage and potential merging of two records at a duplicated site to create a unified record. The de-duplication step was generally undertaken between the GRDB and a 'paired' dataset (e.g. GRDB and GAME). The only exceptions for this step are for GRDB, EWA and ARCTICNET, as these three datasets share Russia as a common spatial domain.

The techniques adopted for combining research databases were based on the de-duplication procedures developed in Gudmundsson and Seneviratne (2016), which consists of three sequential steps:

(1) **Identification of 'duplication-candidates' using metadata similarity.** This step aims to identify timeseries with a high level of similarity in

metadata (either within one database or across different databases). We used three similarity metrics to identify potential timeseries: (1) Jaro–Winkler distances, a metric representing the alphanumeric similarity of strings (Christen, 2012), applied to river names of two records; (2) Jaro–Winkler distances between station names of two records; and (3) geographical proximity estimated from geographical coordinates between two records. These metrics were normalised to have the same range between 0 and 1, where a value of 0 indicates identical metadata (e.g. the same geographic coordinates). This similarity analysis was run for each pair in the pool of stations, and any pair with an average value below 0.25 was identified as candidate duplicate records.

(2) **Classifications of duplication candidates using timeseries similarity.** This step aims to decide whether a specific pair of duplication candidates is likely to be identical. The overlapping period and correlation coefficient were used as criteria for making a decision. Firstly, all duplication candidates that do not share any overlap in their period of record are kept in the final GSIM collection, as they can represent separate timeseries even if they measured discharge at the same geographical location (e.g. due to reconstruction of the gauging station). Secondly, any timeseries with a correlation coefficient (R^2) lower than 0.90 was automatically identified as 'very likely different' (26 pairs), whereas $R^2 > 0.99$ indicates 'very likely identical' timeseries (786 pairs). Finally, candidates with $0.90 \leq R^2 \leq 0.99$ (65 pairs) were visually inspected and manually classified as 'very likely identical' (60 pairs) or 'very likely different' (five pairs). All timeseries in the 'very likely different' category were retained while stations of the 'very likely identical' category were processed using the de-duplication procedure (see below).

(3) **De-duplication of identical timeseries:** regardless identical timeseries come from either the same database or from different databases, records with the greater number of data points in the streamflow timeseries were kept while the other(s) were discarded. Although this approach has the downside of truncating the length of useful records, the number of

timeseries that could be influenced by this approach is relatively low (846 timeseries, corresponding to 2.8% of the total number of available streamflow records).

A visual example of the de-duplication procedure is provided in Figure 2-2. The left panel demonstrates a case of 'very likely identical' stations, when station number 2964035 in the GRDB database was identified as an identical gauge to W.16 in the GAME archive, based on the similarities between the provided metadata and correlation coefficient. The timeseries representing station 'GAME_W.16' was kept in the final collection, while timeseries 'GRDB_2964035' was removed. The right panel in Figure 2-2 demonstrates a case of a 'duplication candidate' with correlation coefficient of 0.92 (timeseries 'GRDB_6123645' and 'EWA_9110028'). These timeseries were visually inspected, assigned a 'very likely different' label, and both timeseries were kept in the final collection.

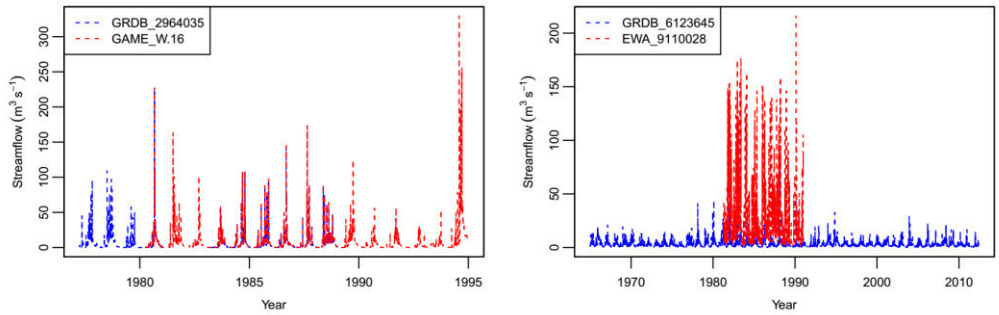


Figure 2-2: Examples of visually inspected duplication-candidate timeseries. Left: Two stations that were labelled 'very likely identical' stations. Right: Two stations that were labelled 'very likely different' stations.

2.4 Production of the GSIM metadata

Providing a consistent set of metadata for each site has been a significant undertaking for GSIM. This section outlines three main stages to developing the GSIM metadata: (1) consolidating all available basic metadata; (2) consistently delineating catchment boundaries for each site; and (3) developing a supplementary set of catchment-scale metadata based on the delineated boundaries.

2.4.1 Consolidating basic metadata from available sources

Following the GRDB format, each timeseries was accompanied by basic metadata, including:

(1) station ID

(2) station name

(3) river name of gauging location

(4) geographical coordinates of station

(5) elevation of station

(6) drainage area

(7) catchment boundary from original data sources.

This data is useful for filtering stations according to specific criteria and analysis objectives. Moreover, the availability of a catchment boundary for the gauge enables additional catchment-scale metadata to be derived as necessary. However, not all of this basic metadata was available for all data sources. For example, the catchment boundary was only available for parts of the GRDB and EWA stations, the drainage area was unavailable in the BOM and MLIT databases, and though several data sources included river names in station names (BOM, HYDAT, USGS), this metadata was unavailable in English for other sources (MLIT, ANA, ADF). Table 2-3 further outlines availability of basic metadata for each source.

Table 2-3: Basic metadata available from data sources

Database	Station ID	Station name	River name	Geo. coor.	Station elevation	Drainage area	Catchment boundary
GRDB	X	X	X	X	X	X	X
EWA	X	X	X	X	X	X	X
CHDP	X	X	X	X	-	X	-
GAME	X	X	X	X	X	X	-
ARCTICNET	X	X	X	X	X	X	-
USGS	X	X	-	X	X	X	-
HYDAT	X	X	-	X	-	X	-
ANA	X	E	E	X	X	X	-
ADF	X	E	E	X	X	X	-
MLIT	X	E	E	X	-	-	-
BOM	X	X	-	X	-	-	-
WRIS	X	X	X	X	X	X	-

(X: metadata available; -: metadata is unavailable; E: metadata is not available in English)

The method for consolidating basic metadata for each station follows three steps:

Step 1. Transfer and review metadata available from original sources.

The transfer of all existing metadata required a range of simple consistency checks and conforming rules, including:

(1) Reviewing geographical coordinates of all stations. Stations with unreasonable locations (e.g. located in the middle of North Atlantic Ocean without any land mass, identified from Google Earth) were marked to be excluded from the subsequent delineation procedure (24 stations).

(2) Separating the river name from the station name. Several sources use a consistent format for the station name consisting of two parts: the name of the station followed by the name of the water body. This pattern used a formula with 'linking words' such as 'at', 'upstream' and 'downstream'. Taking station 'BOM_406219' with original station name 'Campaspe River at Lake Eppalock (Head Gauge)' as an example, the position of linking word 'at' was identified and used to extract 'river' metadata (Campaspe River) from the full station name.

(3) Retaining the metadata of duplicated timeseries having the most data points in contrast to the other timeseries being removed. While this step may mistakenly remove some information, it is expedient and reflects the typical result of de-duplicated records that longer timeseries were kept while the shorter timeseries were removed.

Step 2. Generate 'database-merging' information

This step documents a summary of efforts taken in creating a consistent set of GSIM metadata, and allows a user to check steps that were taken or to identify better procedures using alternative timeseries or metadata obtained from original sources. There are 12 fields documented for this purpose:

(1) an indication of whether the timeseries de-duplication procedure was used (one field)

(2) which database and station was kept to construct the GSIM timeseries (two fields)

(3) which station that were removed and the corresponding database (three fields)

(4) the value of metrics that represent similarities in the timeseries metadata (five fields)

(5) the number of overlapping days, if applicable (one field)

Step 3. Generate information about data availability

The last step in compiling basic metadata for GSIM was to generate metrics that represent data availability for each GSIM timeseries, including the temporal coverage (i.e. the first and final year), the number of available daily observations, the number of missing data points, and the proportion of missing data points.

2.4.2 Catchment delineation procedure

With the ever-increasing availability of remote-sensing and modelled data products at global and continental scales, the provision of catchment boundaries is an important mechanism for extending the utility of GSIM. Although catchment boundaries can be generated easily using standard delineation algorithms in GIS packages, it requires a global coverage DEM dataset and reliable location to represent the outlet of each drainage area, which were unfortunately not readily available for GSIM project. This section describes the DEM products, and the algorithm to identify the 'best outlet location' associated with each station that has been used in GSIM project.

The main DEM product used for GSIM was HydroSHEDS (http://hydrosheds.org), which is available at 15 arc-second resolutions (Lehner et al., 2006), and has been used extensively in large-scale hydrological studies (Lehner et al., 2008, Lehner and Grill, 2013, Wood et al., 2011, Do et al., 2017). To address a limitation in the coverage of HydroSHEDS (no information in regions above 60 degrees North, and some islands), the Viewfinder Panoramas elevation product at 15 arc-second resolutions was used (http://viewfinderpanoramas.org, last accessed 25 June 2017) for those locations. This dataset has been used in several studies as an alternative DEM

product to overcome similar data coverage issues (Sil and Sitharam, 2016, Fredin et al., 2012, Barr and Clark, 2012, Yamazaki et al., 2015). As there were more than 30,000 stations needing to be delineated, the HydroBASINS dataset was used, dividing the world into 24 regions, so that the task of delineation could be performed in parallel. The regions are shown in Figure 2-3 and are generally independent in terms of drainage areas (Lehner and Grill, 2013). North America and Europe were specifically broken into more regions to address their relatively higher density of gauges. To maintain consistency when delineating boundaries, only one DEM product was used per GSIM region. As the quality of the Viewfinder Panoramas is not as clearly documented as for HydroSHEDS, its use was kept to a minimum. This resulted in five regions using Viewfinder DEM and 19 regions using HydroSHEDS (see Table 2-4).

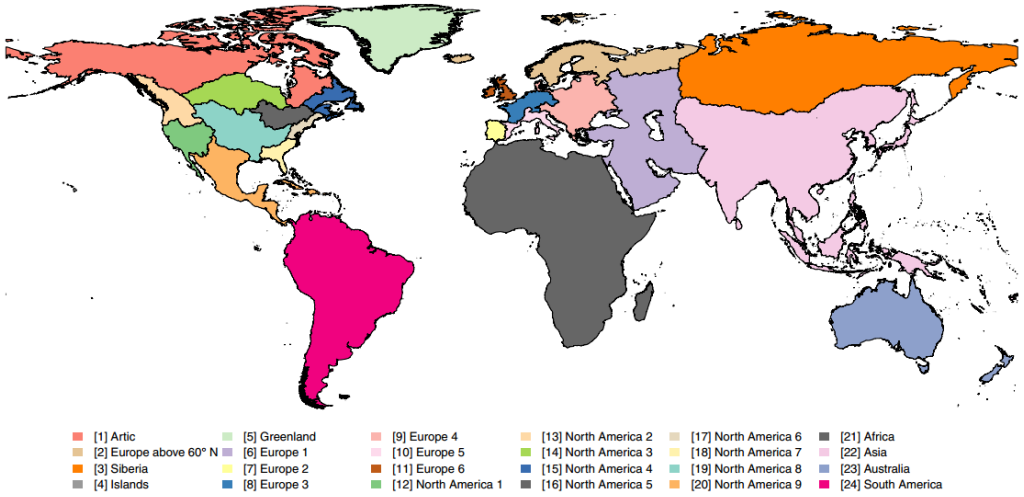


Figure 2-3: GSIM regions for catchment delineation and metadata extraction procedures.

Other challenges in the catchment delineation procedure are possible errors in the geographical coordinates represent the catchment outlet, such as typos in reported coordinates (e.g. 13.47N instead of 14.47N), or swapped order of the coordinate digits (e.g. 103.45E instead of 103.54E). These errors can lead to unreliable results of the delineation procedure, and so that an algorithm to identify location that well represents catchment outlet was also applied. This is described below.

Case 1. Reported station coordinates adopted as the outlet

If there was no information about a drainage area in the station metadata, the geographical coordinates of the station available from the data source were used

as the outlet of the delineation process. There are automated techniques for repositioning outlets, such as choosing cells with the greatest flow accumulation within a search-distance (Snap Pour Point ArcGIS tool), or finding the nearest cell possessing a flow-accumulation value above a specified threshold (Lindsay et al., 2008). Nonetheless, without information on the catchment area, it is impossible to assess the quality of the delineated catchment. Even if a repositioning technique were adopted, delineated catchment boundaries should be used with caution in this case, and therefore the original geographical coordinates was used to represent 'best outlet location'.

Table 2-4: DEM products used for each GSIM region

Region	Description	DEM product
Artic (region 1)	Represents the distant part of North America (including Alaska, most part of Canada and eastern part of Autonomous Province, Russia)	Viewfinder DEM 15s
Europe above 60N (region 2)	Represents countries located above 60°N (e.g. Sweden, Denmark, Norway, part of Germany, part of Russia)	Viewfinder DEM 15s
Siberia (region 3)	Represents areas above the 60°N part of Asia	Viewfinder DEM 15s
Islands (region 4)	Represents some islands across Pacific Ocean (e.g. Honolulu, U.S.) and Atlantic Ocean	Viewfinder DEM 15s
Greenland (region 5)	Represents land mass of Greenland	Viewfinder DEM 15s
Europe 1 to Europe 6 (six regions, from region 6 to region 11)	Represent most European countries (below 60°N)	HydroSHEDS DEM 15s
North America 1 to North America 9 (nine regions, from region 12 to region 20)	Represent U.S. (except Alaska) and the southern part of Canada (below 60°N). It also includes Central America for simplicity in processing catchment boundaries.	HydroSHEDS DEM 15s
Africa (region 21)	Represents Africa region	HydroSHEDS DEM 15s
Asia (region 22)	Represents Asia region (part of Kazakhstan, China, Mongolia and Russia)	HydroSHEDS DEM 15s
Australia (region 23	Represents Australia, New Zealand and some pacific islands	HydroSHEDS DEM 15s
South America (region 24)	Represents South America	HydroSHEDS DEM 15s

Case 2. Application of an automated repositioning algorithm

For stations with available information on catchment area, the automated repositioning procedure documented in GRDC report number 41 (Lehner, 2012) was used with some minor adjustments, and is summarised below:

(1) The catchment area was estimated using the flow accumulation dataset derived from the DEM products. This calculation was repeated for all pixels of the HydroSHEDS/Viewfinder gridded river network within a search radius of 5 km from the geographical coordinates of a specific station.

(2) The estimated area values were compared with the reported area in the original metadata. All pixels were coded with the absolute value of their area differences (in %, with reported area in the metadata used as reference). Pixels with area differences of more than 50% were excluded. This procedure provided an area-based ranking scheme (*RA*) ranging from 0 to 50, where 0 indicates perfect agreement in catchment areas.

(3) The distance to the original location of the station (geographical coordinates reported in the original metadata) was calculated for each pixel and normalised to reach 50 at the maximum distance of 5 km. This procedure provided a distance-based ranking scheme (*RD*) ranging from 0 to 50, where 0 indicates perfect agreement in station locations.

(4) The final ranking scheme (*R*) was calculated as a combination of *RA* and *RD*, where distance rank was weighted twice as high ($R = RA + 2RD$) to penalise pixels that were further away from the original location.

(5) The outlet was automatically relocated to the position of the pixel showing the lowest ranking value, and geographical coordinates of the pixel centroid were defined as the 'best' outlet for this specific catchment.

(6) In the original technical document (Lehner, 2012), a manual procedure was adopted for stations with differences in area above 50 (i.e. the search algorithm cannot find any pixel with an area difference less than 50% within the 5 km search radius), or for stations that had no reported area in the data catalogue. This manual inspection process was infeasible given the scope of the GSIM project, having over 30,000 catchments being delineated and where river names were not available (or potentially inaccurately translated) for many stations.

A Python script was developed to automatically call the 'best outlet location' algorithm and the catchment delineation toolset available in ArcGIS software (Jenson and Domingue, 1988) for each gauge using the chosen DEM data-product. The delineated catchment boundary for each station was assigned a quality flag according to discrepancy between reported drainage area and

delineated catchment boundary area. There are four quality categories associated with the catchment boundary:

(1) 'High' quality: Area difference less than 5%

(2) 'Medium' quality: Area difference from 5% to less than 10%

(3) 'Low' quality: Area difference from 10% to less than 50%

(4) 'Caution' quality: Area difference greater than or equal to 50%, or the reported catchment area was not available in the GSIM catalogue.

Figure 2-4 demonstrates an example where the repositioning algorithm was used. Here the 'best outlet location' was determined to be 4.8611 km away from the original location, which is defined by the reported geographical coordinates in the metadata (for station AR_0000007). The reported area in the metadata is 340km^2 while the area of delineated catchment boundary using original coordinates was only 0.8km^2, which is significantly lower than the correct number. On the other hand, the delineated catchment boundary using 'best outlet location' has an area of 363km^2, indicating a better estimation of the upstream catchment boundary for this particular station.

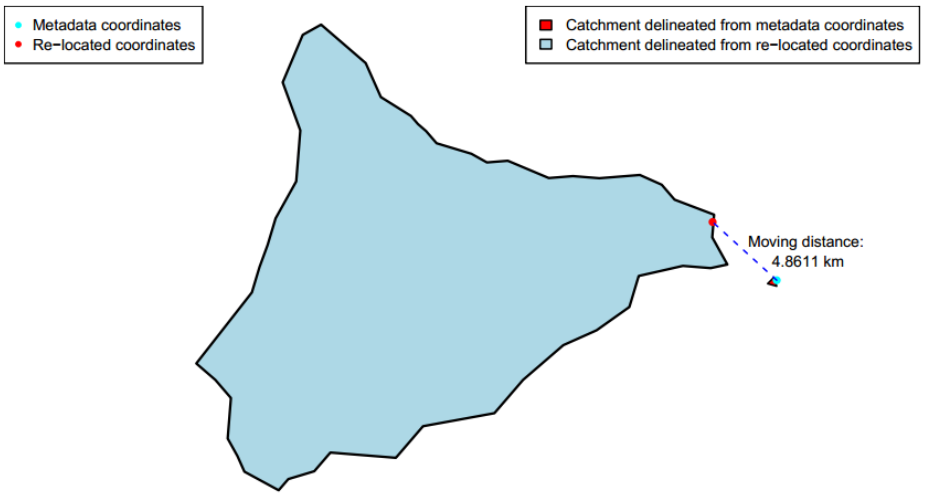


Figure 2-4: Example of improvement in quality of catchment boundary using re-located geographical coordinates (for station AR_0000007).

2.4.3 Extraction of catchment-scale metadata

An important aspect of large-scale hydrology is the ability to exploit gridded datasets at the global scale (Bierkens, 2015, Bierkens et al., 2015, Ward et al.,

2015, Gudmundsson and Seneviratne, 2015, Seneviratne et al., 2012). Having developed catchment boundaries for each GSIM station enabled a supplementary set of catchment-scale metadata to be derived with relative ease. A key feature is that the catchment boundaries and the subsequent metadata relates to the upstream contributing area that influences a gauge, rather than to the catchment (or arbitrarily defined sub-catchment) that contains the gauge and therefore includes a non-influencing downstream region.

In developing the catchment-scale metadata, a standard set of variables have been identified with a view to supporting a range of applications such as filtering stations according to characteristic features, performing analyses of streamflow according to explanatory features of a catchment, or classifying stations according to the (in)significance of human impact. As summarised in Table 2-5, a total of 12 global data products were used to derive 19 elements of catchment-scale metadata. These products were chosen to represent five main categories of catchment characteristics: (1) topography, (2) human impact, (3) climate type, (4) vegetation type, and (5) soil profile. Because the global data products have varying resolution and structure, the following method was used to derive the catchment-scale metadata:

(1) Delineated catchment boundaries associated with each stream gauge were used to mask the subset of pixels from the resampled dataset.

(2) If more than 30% of the catchment area was not covered by a specific global data product, a 'No data' code was given.

(3) Metadata representing the characteristics of the upstream catchment for each streamflow gauge were calculated from the gridded data masked in step (1). There were three types of metrics calculated during this step:

   (a) **A single value**. Used only for the elevation at the geographical coordinates of the gauge (i.e. the catchment outlet), number of large dams located within the catchment boundary, and total volume of corresponding reservoir.

   (b) **Average, min, max and quartiles values.** Used for continuously varying data such as a slope or topography index. These metrics

allow an idea of central tendency as well as spread of extracted data within each catchment boundary.

(c) **Percentages of different classes of catchment characteristics.** Used for categorical data. For example, there are 16 classes in the global lithology dataset, and the co-presence of more than one type of lithology occurs very often across all catchments. The percentages of each lithology class were therefore calculated and recorded for all available catchments. To make the results presentable in a final catchment-scale metadata matrix, an aggregated metric was calculated to indicate that there is a dominant class within the catchment boundary (i.e. more than 50% of all available pixels). If there is no dominant class within the catchment boundary, a 'No dominant class' string is provided.

Table 2-5: Global data-products used in GSIM and derived catchment-scale metadata

Variables	Data sources	Spatial Resolution	Reference period	Extracted metadata
Elevation	HydroSHEDS http://hydrosheds.org/ ViewFinder http://viewfinderpanoramas.org/	15 arc-seconds x 15 arc-seconds	-	(1) Gauge elevation (2a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of catchment elevation
Slope	Derived from HydroSHEDS and ViewFinder D.E.M by authors	15 arc-seconds x 15 arc-seconds	-	(3a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of catchment slope
Topographic index	High-resolution global topographic index values (Marthews et al., 2015) https://catalogue.ceh.ac.uk/documents/ce391488-1b3c-4f82-9289-4beb8b8aa7da	15 arc-seconds x 15 arc-seconds	-	(4a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of catchment topographic index
Drainage density	GRIN - Global River Network (Schneider et al., 2017) https://www.metis.upmc.fr/fr/node/375	7.5 arc-minutes x 7.5 arc-minutes	-	(5a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of catchment drainage density (km^{-1})
Dams	Global Reservoir and Dam (GRanD), version 1 (Lehner et al., 2011) http://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev01	6,862 datapoints storage capacity of more than 0.1 km3	-	(6) Number of dams upstream (7) Total upstream storage volume
Population	Gridded Population of the World (GPW) version 4 (CIESIN, 2016) http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count	30 arc-seconds x 30 arc-seconds	2005-2014	(8a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of catchment population (2010) (9) 2010 Population count
Urbanisation	Night Light Development Index (NLDI) dataset (Elvidge et al., 2012) http://www.soc-geogr.net/7/23/2012/sg-7-23-2012.html	0.25 arc-degrees x 0.25 arc-degrees	2006	(10a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of NLDI over catchment
Irrigation	Historical Irrigation Dataset (Siebert et al., 2015) https://mygeohub.org/publications/8/2	5 arc-minutes x 5 arc-minutes	2005	(11a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of catchment Irrigated area (2005)

Variables	Data sources	Spatial Resolution	Reference period	Extracted metadata
Climate type	World map of Koppen Weiger climate classification system (Rubel and Kottek, 2010) http://koeppen-geiger.vu-wien.ac.at	5 arc-minutes x 5 arc-minutes	1951-2000	(12) Type of catchment climate (Koppen-Weiger) if one type present over more than 50% catchment area, or 'No dominant type'
Land cover	The Climate Change Initiative Land Cover (CCI-LC) dataset http://maps.elie.ucl.ac.be/CCI/viewer/download.php	7.5 arc-seconds x 7.5 arc-seconds	2015	(13) Type of catchment land-cover (UN Land Cover Classification System) for 2015 if one type present over more than 50% catchment area, or 'No dominant type'
Lithological	The Global Lithological Map v1.0 (GLiM) dataset (Hartmann and Moosdorf, 2012) https://www.clisap.de/research/b:-climate-manifestations-and-impacts/crg-chemistry-of-natural-aqueous-solutions/global-lithological-map/	0.5 arc-degrees x 0.5 arc-degrees	-	(14) Type of catchment lithology if one type present over more than 50% catchment area or 'No dominant type'
Soil profile	Soil grid 250m (Hengl et al., 2017) https://soilgrids.org	7.5 arc-seconds x 7.5 arc-seconds	-	(15) Type of catchment soil-class (World Reference Base) if one type present over more than 50% catchment area or multiple types 'No dominant type'. (16a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of weight percentage of sand over the catchment (17a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of weight percentage of silt over the catchment (18a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of weight percentage of clay over the catchment (19a-f) Average, minimum, maximum, first quartile, second quartile and third quartile values of bulk content of soil over the catchment (kg/m^3)

2.5 Overview of the GSIM archive

This section summarises the GSIM archive, including the availability of timeseries combined from 12 original data sources, the associated data-products and documentation outlining data quality (section 2.5.1). The whole time series database cannot be made available online due to data policies from a number of original data sources, some of which apply very strict terms and conditions regarding to redistribution of streamflow time series. To address this limitation and maintain the usefulness of GSIM to the research community, three metadata products have been developed and the availability of these data products is further discussed in section 2.5.2.

2.5.1 Timeseries availability

From the 35,002 timeseries records obtained from 12 different sources, the final GSIM timeseries archive holds a total of 30,959 unique stations, of which 30,935 stations have associated catchment shapefiles and catchment-scale metadata (24 stations were removed from this process due to suspect geographical locations). Most data sources are still active and being updated by the data authorities. GSIM, however, also included 425 "static" time series (from ARCTICNET, GAME and CHDP databases) that have been frozen since the early 2000s as these stations have improved the gauge density in regions with sparse streamflow observation systems (Russia, China and Thailand respectively). In addition, 2,735 EWA stations (frozen since October 2014) were also included into GSIM as these timeseries have not been completely mirrored into GRDB database at the time GSIM was initiated. As these "static" time series have been frozen and no further update were provided, GSIM users are advised to use them with caution as the data may contain errors and/or have been replaced or updated.

As shown in Table 2-6, it is apparent that spatial coverage of the stations in the GSIM database varies significantly across continents, with North America and Europe having the greatest number of stations. Including the national databases such as MLIT (Japan), ANA (Brazil), BOM (Australia), and IWRIS (India) has significantly improved the observational network over the regions of Asia, South America and Oceania (top panel of Figure 2-5),

some of which have recorded streamflow since the mid-20th century and were still operating at the time the GSIM database was initiated. This suggests that the national databases that are currently available should be given more attention in order to improve the quality and quantity of international archives.

Table 2-6: Summary statistics of GSIM timeseries

Continent	Number of stations	Average temporal coverage (years)	Shortest record (years)	Longest record (years)	Year of earliest entry	Year of latest entry
Africa	949	33.8	1	110	1903	2015
Europe	5,778	40.3	1	208	1806	2016
Asia	1,915	22.2	1	79	1921	2015
North America	15,884	42.9	1	156	1860	2016
South America	3,449	29.3	1	116	1901	2016
Australia and Oceania	2,984	31.4	1	131	1886	2016
Global	30,959	38.2	1	208	1806	2016

Regarding temporal coverage, streamflow records across the globe are generally available for the second half of the 20th century (as shown in the bottom panel of Figure 2-5). Regardless of missing data criteria, the number of available data gradually rises to its peak in the late 1970s to early 1980s, followed by a mild decrease in the late 1980s as also discussed by Hannah et al. (2011) and a secondary peak in the late 2000s. While the overall database has over 30,000 gauges, it is clear from Figure 2-5 that from the 1960s onwards there are approximately from 10,000 to 15,000 gauges simultaneously active. This represents a significant increase in availability compared to the GRDB dataset, which had a total of approximately 9000 gauges and with a similar drop-off in available gauges depending on the filtering criteria applied.

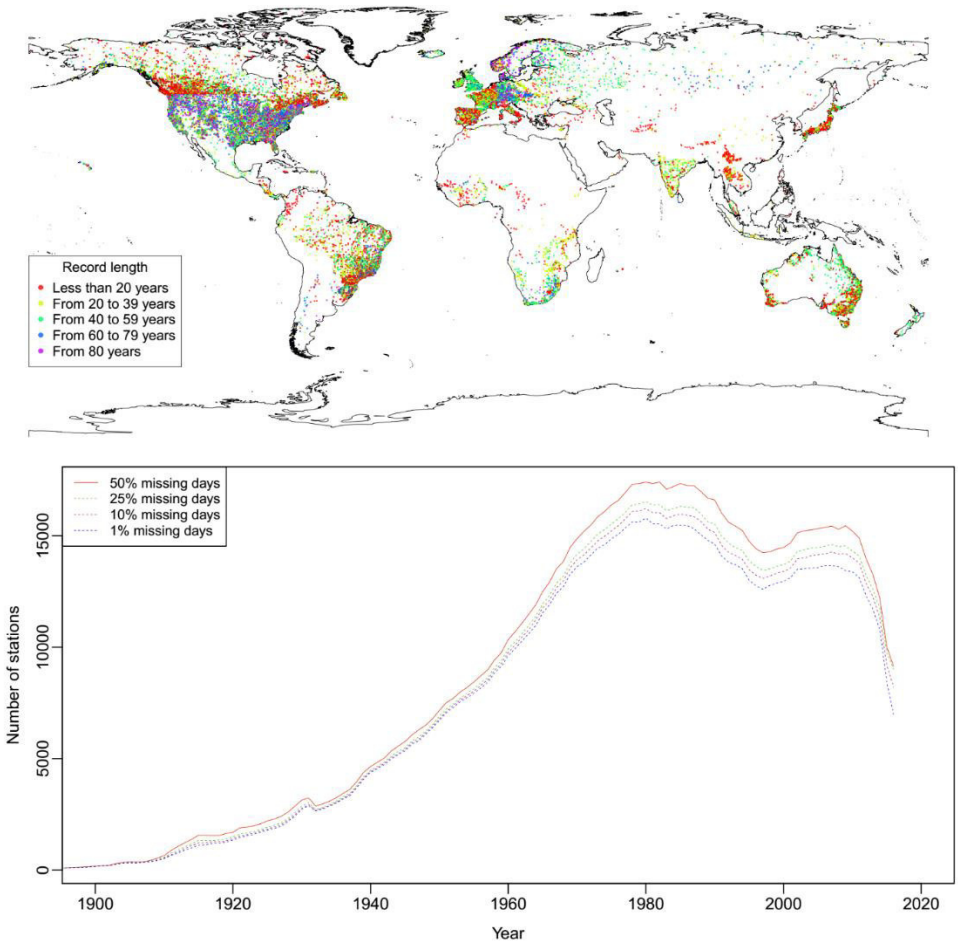


Figure 2-5: Availability of GSIM timeseries. The top panel illustrates the length of record at each station, and the bottom panel illustrates the number of available time series over time for four different missing data criteria.

2.5.2 Data products of GSIM

2.5.2.1 GSIM catalogue

The GSIM catalogue is designed for users to easily filter stations according to their purpose of application, and where necessary to transparently identify steps taken in the development of GSIM. The total number of 27 fields included in this document can be divided into three groups, namely:

(1) Basic metadata: This group provides station identification, including a unique GSIM number, the name of the river, the name of the station, the elevation of the gauge, the provided geographical coordinates, and the catchment area.

(2) Database merging metadata: This group of fields provides the identity of the numbers of original source(s), and if applicable the similarity metrics between duplicates.

(3) Data availability metadata: This group of fields provides an overview on the data availability of each timeseries. These statistics were generated from the timeseries data and can be used to filter station information, such as temporal coverage, data length, and the fraction of missing data.

As illustrated in Table 2-7, source datasets had significant gaps in the metadata, especially in cases of gauge elevation (not available in CHDP, GAME, HYDAT, BOM, MLIT) and catchment area (not available in BOM, MLIT). In addition, the geographical coordinates of all stations were not correctly recorded for all stations, with 24 removed as having suspect locations and 4,871 shifted coordinates as part of the procedure for aligning catchment outlets with reported catchment areas.

Table 2-7: The percentage of stations accompanied by all basic metadata

Dataset	Station ID	River name	Station name	Latitude	Longitude	Altitude	Catchment area
ADF	100	100	100	100	100	96.2	99.3
ANA	100	99.9	100	100	100	69	99
ARCTICNET	100	100	100	99.3	99.3	99.3	100
BOM	100	100	100	100	100	0	0
CHDP	100	99.4	100	100	100	0	84
EWA	100	100	100	100	100	98.5	94.5
GAME	100	100	100	100	100	0	100
GRDB	100	100	100	100	100	67	100
HYDAT	100	100	100	100	100	0	85.8
MLIT	100	100	100	100	100	0	0
USGS	100	100	100	100	100	93.7	25.5
WRIS	100	100	100	100	100	81.6	97.4
GSIM	100	99.9	100	99.9	99.9	50.4	73.8

2.5.2.2 Quality of catchment boundary

The catchment boundary is the second metadata product that is available through GSIM. Of all GSIM stations, 12,150 (39%) were not associated with any information about drainage areas (including all MLIT and BOM stations); thus, a ‘Caution’ flag is attached to upstream catchments of these stations. Another 24 stations with suspected geographical coordinates of stations were

removed, and the final 18,785 stations were processed to identify the 'best outlet' location to represent the outlet for delineating upstream catchments. The distribution and quality of the delineated catchments of these stations are provided in Figure 2-6 (figures at continental scale are also provided as supplementary materials).

As illustrated in the top panel, 'Caution' catchments using 'best' outlets (identified using the method outlined in Section 2.4.2) are generally located across all GSIM regions. However, the 'Caution' flag appears more frequently over regions above 60 degrees North. Further checks would be required to improve the association of catchment boundaries to stations. Unfortunately, the biggest caveat that applies to the GSIM database, as with any global database, is that the metadata were collated from a number of sources with varying standards of documentation and quality assurance and with limited capacity for additional checking other than automated procedures. Therefore, there is likely to be a non-trivial degree of error in the metadata for both geographical location and drainage area. Another issue that may lead to unreliable results of the delineation process is error in the DEM products. This potential error has been documented (Lehner et al., 2006, Lehner, 2012), and lower quality DEM products generally exist for regions above 60 degrees North due to the lower quality of the original elevation products used to derive the DEM datasets. Another note for the use of delineated catchments is that very small catchments (area less than 50 km^2) should be handled with care, as the 'best' outlets could be located incorrectly while still delivering 'acceptable' discrepancies as part of the automated procedure.

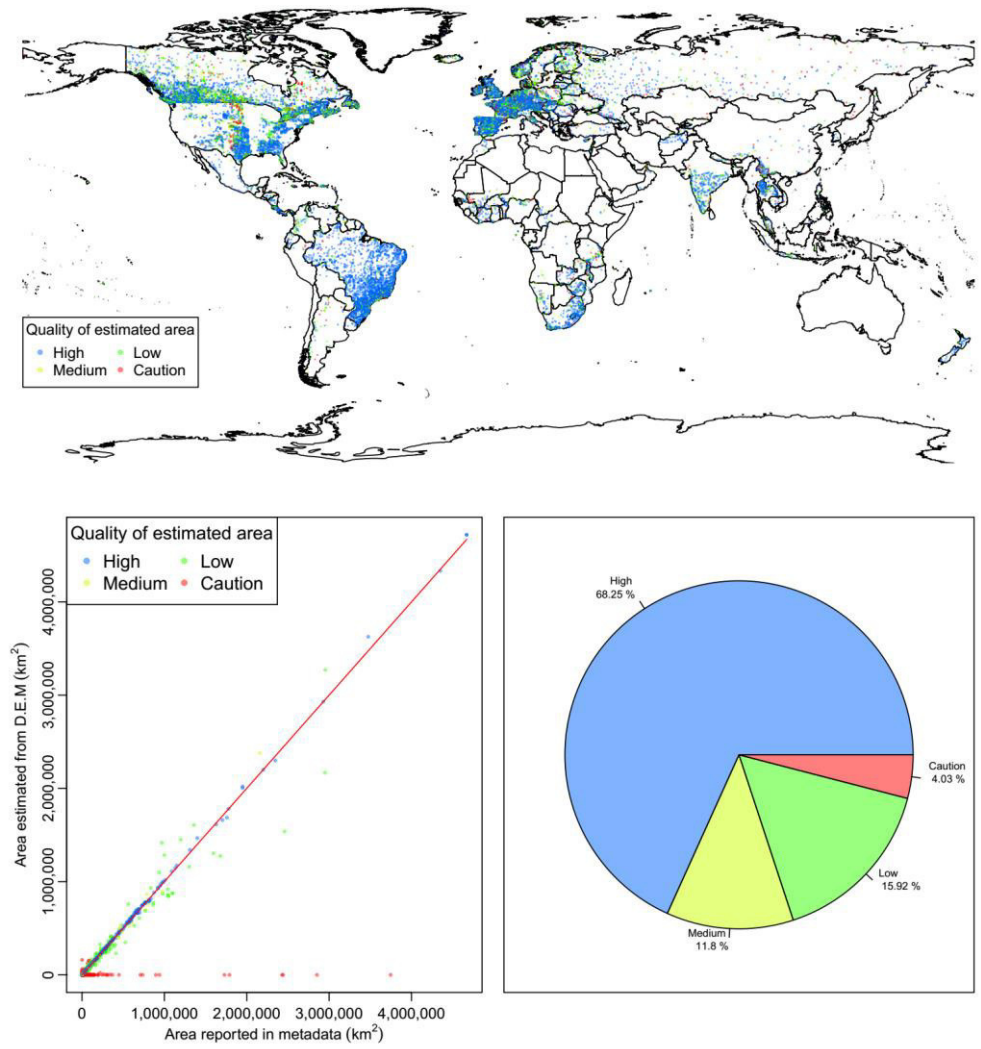


Figure 2-6: Quality of delineated catchment boundary according to categories of high, medium low and caution identified in Section 2.4.2 (for 18,785 stations that have reported drainage area and reasonable geographical coordinates).

Nonetheless, the quality of delineated catchments is quite positive (as illustrated in lower panels of Figure 2-6). Of all 18,785 catchments that had reported drainage area in the GSIM catalogue, 68.25%, 11.8% and 15.92% catchments have 'high' quality (area discrepancy of less than 5%), 'medium' quality (area discrepancy from 5% to less than 10%) and 'low' quality (area discrepancy from 10 to less than 50%) respectively, while there are only 4.03% catchments with 'Caution' quality (area discrepancy of more than or equal to 50%).

2.5.2.3 Catchment-scale characteristics

The final data product that has been made available is the auxiliary information extracted from 12 global coverage datasets representing many

characteristics associated with GSIM stations. Overall, the spatial coverage of original data products (mostly satellite-base) are quite good (see Table 2-8), with just a small fraction of catchments (less than 10%) that have more than 30% of their areas not covered by these datasets. The exception is the Nightlight Development Index (NLDI – computed from the 2006 Nightlights dataset (Ziskin et al., 2010) and the 2006 Landscan gridded population (Bhaduri et al., 2002)). This dataset does not have approximately 25.3% of catchments covered, for more than 70% of their areas.

Table 2-8: Percentages of available catchment-scale characteristics

Catchment characteristics	Number of stations	Availability percentage
Climate classification	30,773	99.5
Drainage density	29,574	95.6
Elevation	30,932	99.9
Irrigation area	30,857	99.7
Land cover classification	30,888	99.8
Lithology type	30,154	97.5
Nightlight Development Index	23,096	74.7
Population count	30,894	99.9
Population density	30,800	99.6
Slope	30,862	99.8
Soil bulk density	30,812	99.6
Soil classification	30,764	99.4
Clay content	30,768	99.5
Clay content	30,695	99.2
Silt content	30,828	99.7
Topographic index	30,725	99.3

It is important to note that while these catchment-scale characteristics are consistent products available for all stations, documentation for the original source data should be consulted during application to appreciate the limitations and appropriateness of each variable. For examples, the GRanD database is not exhaustive of all dams worldwide and there can be ambiguities over the affiliated dates (e.g. whether they represent conception, construction or commissioning). Furthermore, the extent of overlapping period between temporal coverage of streamflow time series and remote sensing based datasets need to be carefully assessed in cause-effect studies. Similarly, it is likely that there will be updated or new data gridded datasets available over time so that applications should consider the appropriateness of the information used. The availability of metadata products emerging from the

GSIM project demonstrates the possibility of using reported global data products to extract catchment-scale characteristics associated with each station with reasonable quality, enabling many potential applications from this rich information.

2.6 Data availability

The data described in this paper are available as a compressed zip-archive containing (i) a readme file, (ii) metadata of all GSIM stations obtained from original data sources and time series, (iii) quality of catchment boundary and catchment characteristics extracted from 12 global data-products, (iv) list of stations with suspect geographical coordinates and (v) catchment boundaries for 30,935 stations that have reasonable geographical location.

The data can be freely downloaded at https://doi.pangaea.de/10.1594/PANGAEA.887477. The uploaded zip-archive containing two directories and one README.txt file. The readme file provides a detailed description of the data. The directory "GSIM_catalogue" contains the the metadata of all GSIM stations and a list of stations with suspect geographical coordinates. The directory "GSIM_catchments" contains shapefiles for 30,935 stations.

2.7 Conclusion

In-situ observations of daily streamflow with global coverage are crucial to understanding large-scale freshwater resources that are fundamental for societal development. The GSIM archive, designed as an expansion of the GRDB database, has demonstrated the possibility of significantly improving the coverage and density of the global streamflow observational datasets using free-to-access databases. The development of the GSIM database was not possible without the tremendous investment into the production and ongoing maintenance of original data sources of GSIM. This fact emphasises the key role of data authorities and international initiatives in enabling advances in large-scale hydrology by making data publicly available to the community.

While the activities of GSIM have been extensive in searching out and collating databases, they are by no-means exhaustive (e.g. since submission we have been notified of additional potential candidates for inclusion such as the Mekong River Commission database, Chile national water database and Argentina national water database). It is the authors' intention that this project will stimulate further efforts toward the development of coordinated and consistent representation of global streamflow observations. For this reason, the process of developing the archive was designed with automation in mind. With the exception of needing to visually inspect some cases of duplicated timeseries, the archive was automated using scripts in the R and Python programming languages.

Although the GSIM database was compiled from data sources that can be obtained free of charge via a data portal or by submitting written requests to data authorities, there are some strict conditions related to the redistribution of un-processed data. Therefore, it is impossible to make the whole GSIM collection publicly available. In addition, with the main aim of harvesting as much data as possible, the GSIM database is not focused on collecting high-quality datasets such as referenced hydrological networks that are available in many countries (Whitfield et al., 2012), and thus the data quality may vary significantly across the available timeseries. To address these limitations and increase the usefulness of the GSIM database, we conducted a set of quality checking procedures for all GSIM timeseries. These quality-assured records were then used to produce a dedicated set of indices capturing important aspects of the daily dynamics from GSIM timeseries, and to explore potential applications of GSIM in large-scale hydrology. Detailed information about this work and associated distributed data is described in the second part of our series on GSIM (Gudmundsson et al., in review).

With the GSIM archive and production information made publicly available in a transparent manner, this project serves the broader hydrology community with improved coverage and quality of streamflow information. This project has yielded a significant increase in the availability of streamflow observations through the process of collating readily-accessed online data, and with ongoing efforts there will be opportunities for further extension.

Streamflow observations represent an underutilized resource, in part due to access limitations, but also due to challenges in accounting for human impacts in the observed record. These challenges notwithstanding, ongoing advances in global-scale hydrological models and ever-increasing access to remote-sensed products indicate that wider access to streamflow data has the potential to significantly enhance our knowledge of global water resources.

Acknowledgement

The authors would like to express their appreciation to all the national agencies and institutions that made the streamflow data available for this study. We would like to thank Sonia I. Seneviratne for her discussions and support on the collation of the GSIM archive. Mr Hong Xuan Do receives financial support from the Australia Award Scholarship (AAS). Dr Westra's time was supported by Australian Research Council Discovery project DP150100411. The authors also wish to thank two reviewers for their constructive comments and suggestions.

Chapter 3. A global-scale investigation of trends in annual maximum streamflow (Paper 2)

Statement of Authorship

Title of Paper	A global-scale investigation of trends in annual maximum streamflow
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Do, H.X., Westra, S., Leonard, M. (2017). A global-scale investigation of trends in annual maximum streamflow. Journal of Hydrology, 552, 28-43.

Principal Author

Name of Principal Author (Candidate)	Hong Xuan Do		
Contribution to the Paper	Designed the experiments, conducted statistical analyses, analysed results, wrote manuscript and acted as corresponding author.		
Overall percentage (%)	85		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	07/02/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Seth Westra		
Contribution to the Paper	Suggest scope of study, helped to design experiments and interpret results, provided feedbacks on manuscripts and responses to reviewers.		
Signature		Date	15/02/2019

Name of Co-Author	Michael Leonard		
Contribution to the Paper	Suggest scope of study, helped to develop Rscript to conduct statistical analysis, helped to evaluate and edit the manuscript manuscripts and responses to reviewers.		
Signature		Date	19/2/2019

Abstract

This study investigates the presence of trends in annual maximum daily streamflow data from the Global Runoff Data Centre database, which holds records of 9,213 stations across the globe. The records were divided into three reference datasets representing different compromises between spatial coverage and minimum record length, followed by further filtering based on continent, Köppen-Weiger climate classification, presence of dams, forest cover changes and catchment size. Trends were evaluated using the Mann-Kendall nonparametric trend test at the 10% significance level, combined with a field significance test. The analysis found substantial differences between reference datasets in terms of the specific stations that exhibited significant increasing or decreasing trends, showing the need for careful construction of statistical methods. The results were more consistent at the continental scale, with decreasing trends for a large number of stations in western North America and the data-covered regions of Australia, and increasing trends in parts of Europe, eastern North America, parts of South America and southern Africa. Interestingly, neither the presence of dams nor changes in forest cover had a large effect on the trend results, but the catchment size was important, as catchments exhibiting increasing (decreasing) trends tended to be smaller (larger). Finally, there were more stations with significant decreasing trends than significant increasing trends across all the datasets analysed, indicating that limited evidence exists for the hypothesis that flood hazard is increasing when averaged across the data-covered regions of the globe.

3.1 Introduction

In recent decades, floods have caused nearly half of all weather-related disasters worldwide, and affected more than two billion people (CRED, 2015). The relative importance of flooding as a natural hazard has also increased over this period, whether measured in terms of economic losses (Kundzewicz et al., 2013), reinsurance losses (Mills, 2005) or the number of reported flood events (Munich Re, 2015, Swiss Re, 2015). Improved understanding of the causes of these changes is critical to manage and mitigate future impacts. However, the attribution of observed changes remains unclear, with possible causes including changes to the magnitude or frequency of high flow events (the flood hazard), the number of people or assets potentially affected by flooding (the flood exposure), the magnitude of impacts given a flood exposure (the flood vulnerability) (Kron, 2005, IPCC, 2012), or changes to reporting mechanisms and practices (Peduzzi et al., 2009).

One possible cause of the observed changes to flood impacts is the potential role of anthropogenic climate change. Recently, numerous studies have shown an intensification of extreme precipitation over data-covered land regions globally, and given that extreme precipitation is a leading cause of disastrous flooding (Guha-Sapir, 2014, Guha-Sapir et al., 2015), this intensification might cause an associated increase in flood hazard and thus flood impact. For example, Min et al. (2011) found that 65% of the data-covered areas of the globe exhibited increasing trends for annual maximum rainfall based on a gridded precipitation data product from 1951-1999. Using a point-based data record of 8,326 high-quality land-based stations, Westra et al. (2013) detected increasing trends at nearly two-thirds of stations, and found that the median intensity of extreme precipitation increased in proportion to changes in global mean temperature at a rate of between 5.9% and 7.7% K^{-1}. Lehmann et al. (2015) also found large-scale increasing patterns in extreme precipitation, with 12% more record-breaking rainfall events over 1981–2010. This change was linked to an increasing trend in global temperatures over this time.

The observed intensification of extreme precipitation led the Intergovernmental Panel on Climate Change (IPCC) to find medium confidence for the conclusion that the increasing trends in extreme precipitation and associated discharge implies greater risk of flooding at the regional scale (IPCC, 2014). However, the implications of changes to extreme precipitation on discharge should be assessed with care due to the additional influence of a catchment's antecedent moisture content (i.e. the moisture stored in the catchment's soils, groundwater, lakes and reservoirs prior to the flood-producing rainfall event), which is affected by a catchment's long-term water balance (Johnson et al., 2016) rather than the intensity of individual heavy rainfall events. For example, only a third of discharge above the 99th percentile corresponded to precipitation above the 99th percentile (Ivancic and Shaw, 2015), indicating that the relationship between changes in extreme rainfall intensity and changes in flood hazard are complex and unlikely to be direct. This suggests that trends in extreme precipitation are not likely to be the only climatic factor influencing flood hazard, so that it is not possible to infer the direction and/or magnitude of change in flood hazard from information about changes in extreme precipitation alone.

An alternative approach to understanding changes in flood hazards at the global scale is through the use of large-scale hydrological models (Arnell and Gosling, 2014, Dankers et al., 2014). Studies that have taken this approach generally have found that more increases in flood hazard occurred in land grid cells than decreases, which is consistent with changes in climate variables generated from global climate models. However, high uncertainty remains a challenging issue in simulating streamflow at regional and local scales, as the direction of change is not always consistent across models for some river basins (Dankers et al., 2014). Although on-going efforts are being made to improve model resolution (Wood et al., 2011, Bierkens, 2015) and the representation of hydrologic processes (Clark et al., 2015), these models are still unable to represent many of the complex processes that are involved in the translation of rainfall to runoff at the global scale (Ward et al., 2015, Sood and Smakhtin, 2015, Archfield et al., 2015).

Rather than rely on indirectly inferring changes in flood hazard based on extreme precipitation data or using large-scale hydrological models to understand historical changes, several studies have directly analysed changes in streamflow data. However, challenges to this approach include data quality, availability, and the range of non-climatic factors that may also cause changes in flood hazard. In particular, compared to atmospheric variables such as temperature, pressure and rainfall, streamflow measurements are more susceptible to local (e.g. catchment-scale) anthropogenic influences. For example, land-use change, de-forestation, dams, reservoirs and other effects of urbanisation can all affect flood magnitude (Bradshaw et al., 2007, FitzHugh and Vogel, 2011), as can other hydraulic influences such as regulated water releases, changing channel capacity and/or implementation of flood prevention measures (Hewlett and Hibbert, 1967, Stover and Montgomery, 2001, Slater et al., 2015). There are also many practical challenges in characterising streamflow processes such as the difficulty of identifying subsurface contributions, inaccurate stage-discharge curves, tidal influences in estuarine catchments, and errors when measuring large flow events (Herschy, 1994, Buschman et al., 2009, Di Baldassarre and Claps, 2011, Ghasemizade and Schirmer, 2013). Finally, political sensitivity and costs associated with digitizing records or sharing data (Nelson, 2009, Hannah et al., 2011) can affect the availability of streamflow records, particularly at the continental or global scale.

Arguably as a result of the abovementioned challenges, studies that investigate changes in flood hazard directly based on observed flow data at the global scale are limited (Milly et al., 2002, Kundzewicz et al., 2004). In contrast to the extreme precipitation studies described earlier, these studies have relied on relatively small datasets (using 29 and 195 time series for the Kundzewicz et al. (2004) and Milly et al. (2002) studies, respectively). Furthermore, they did not lead to consistent conclusions regarding changes to flood hazard, with Kundzewicz et al. (2004) finding a similar number of stations exhibiting increasing and decreasing trends, and Milly et al. (2002) finding substantial increases in flood frequency for large basins.

In contrast to the limited number of global-scale studies, there have been numerous studies at regional scales, and some of these have been summarised in Table 3-1. However, it can be difficult to infer a picture at the global scale from these studies due to varying periods of data and strategies for selecting stations. Moreover, little effort has been made in relating any observed changes to natural or anthropogenic factors (Merz et al., 2012b) and there is an ongoing need for studies that use a consistent methodology for assessing streamflow across all regions to advance our knowledge of historical changes in floods (Kundzewicz et al., 2012, Milly et al., 2015, Merz et al., 2012a).

Table 3-1: Summary of observation-based studies on changes to flood frequency and/or magnitude.

Studies	Dataset	Scale of study	Main Findings
Milly et al. (2002)	29 daily time series spanning 1865 – 1999 from GRDC	Global	A substantial increase in the frequency of floods with discharges exceeding 100-year levels was identified from analysing annual maximum monthly-mean flows at 29 basins larger than 200,000 km^2
Kundzewicz et al. (2004)	195 daily time series spanning 1824 – 2002 from GRDC	Global	Annual maximum mean daily streamflows were analysed using the Mann-Kendall test at the 10% significance level. Only 27 (31) gauges had statistically significant increasing (decreasing) trends, and most (137) time series did not show any significant changes
Zhang et al. (2001)	249 Reference Hydrometric Basin Network (RHBN) records covering the 1947 – 1996 period	Canada	Annual maximum streamflows were analysed using the Mann-Kendall test at the 10% significance level, with significant decreasing trends in southern Canada and increasing trends in northern Canada
Cunderlik and Ouarda (2009)	169 stations in the RHBN with common period from 1974 – 2003	Canada	Seasonal maximum streamflows over the snowmelt period were analysed using the Mann-Kendall test at 10% significance level. Almost 20% of all stations showed significant trends in the magnitude of snowmelt floods. A notable finding is most of the detected significant trends were decreasing.
Burn and Whitfield (2016)	280 stations obtained from Environment Canada Data Explorer database spanning the 1961 – 2010 period	Canada	Changes in the magnitude and timing of flood events exceeding the 25-, 50- and 100-year return periods were examined. The study found a generally decreasing flood magnitudes in nival catchments, and increasing flood magnitudes in pluvial catchments.
Lins and Slack (1999)	395 time series covering the 1944 – 1993 period from U.S. Hydro-Climatic Data Network (HCDN)	U.S.	Annual maximum streamflows were analysed using the Mann-Kendall test at the 5% significance level. The study found a mix of increasing and decreasing trends in annual maximum streamflows across the eastern half of the United States, while trends in the

Studies	Dataset	Scale of study	Main Findings
			western U.S. record mostly decreased.
Douglas et al. (2000)	1474 time series spanning over 1874 – 1988 from HCDN	U.S.	Annual maximum daily streamflows were analysed using the Mann-Kendall test at the 5% significance level. Across U.S., no evidence of trends in flood flows was found.
Mallakpour and Villarini (2015)	774 U.S. Geological Survey (USGS) stream gauges covering the common 1962-2011 period	Central U.S.	This study focused on assessing changes of flood magnitude and frequency in the Central United States using the Mann-Kendall test at the 5% significant level. The findings showed that flood frequency has increased while there was limited evidence of a decrease in flood magnitude in this region.
Archfield et al. (2016)	345 USGS time series spanning the 1940–2013 period	U.S.	Four dimensions of floods (frequency, magnitude, duration and volume) were evaluated across U.S. Although detected trends were more than what would be expected by chance, this study could not identify any clear spatial pattern of changes in floods.
Slater and Villarini (2016)	Daily gage height data over 2042 catchments collected from USGS database	U.S.	The peak over threshold approach was used to identify the number of days exceeding the threshold of gage height data. Trends were detected using a Poisson regression model at the 5% significance level. This study identified that there were increases in flood risk around the upper Midwest/Great Lakes region and decreases on the Gulf Coastal Plain, the southeastern United States, and California
Stahl et al. (2011)	441 time series spanning over 1962 – 2004 obtained from European Water Archive (EWA)	Europe	This study calculated the slope of Kendall-Theil robust line for 7-day annual maximum streamflows records and compared it with model simulation from global hydrology model. An increasing pattern in the north/west and decreasing pattern in the south/east of the European continent was found.
Morán-Tejeda et al. (2012)	57 stations spanning the 1961–2005 period obtained from Duero water management agency	Spain	The 90^{th} and 99^{th} percentile of daily streamflow over a year was used to represent flood data. This study showed a general trend of decreasing frequency and magnitude of high flows throughout most of the basin.
Bard et al. (2012)	342 AdaptAlp stations (obtained from seven data authorities) with at least 40 years of data records over the common period 1961–2005	Alps (Europe)	Snowmelt indices representing snowmelt streamflow intensity and seasonality were statistically investigated at the 10% significance level. This study showed an increasing trend in spring floods associated with snowmelt.
Hannaford et al. (2013)	132 time series covering 1932 – 2004 period from EWA	Europe	Trends in 7-day annual maximum streamflow were indicated by the Mann-Kendall Z statistic. The results demonstrated that trends in flood magnitude are highly influenced by interdecadal variability.

Studies	Dataset	Scale of study	Main Findings
Ishak et al. (2013)	330 stations spanning over 1955 – 2004 obtained from the Bureau of Meteorology	Australia	Annual maximum daily streamflows were analysed using the Mann-Kendall test at the 10% significance level. The assessment of trends indicated that the south-east and south-west regions of Australia have experienced a significant downward trend in the annual maximum streamflow over reference periods.
Delgado et al. (2010)	4 stations spanning1913 – 2000 the Southern Institute of Water Resources Research – Vietnam	South-east Asia	Annual maximum daily streamflow were analysed using linear regression, the Mann-Kendall test at 10% significance level and a generalised extreme value model. The study found that the probability of average flood events has decreased during recent decades while extreme floods are likely to increase.
Nka et al. (2015)	11 time series covering 1950 – 2010 period	Africa	Annual maximum daily streamflow were analysed using the Mann-Kendall test at 10% significance level. This study indicated that a mixture of both increasing and decreasing trends were found across West Africa.

In order to better understand historical changes in flood hazard at the global scale, this paper therefore aims to improve the current understanding of changes to annual maximum streamflow by studying changes in the magnitude of annual maximum streamflow from a large global dataset of daily streamflow observations. The dataset is presented in Section 3.2, together with a number of catchment characteristics (climate region, catchment size, de-forestation and presence of large dams) that are used to screen the quality of records and investigate changes in streamflow with respect to potential influencing factors. The annual maximum daily streamflow and analysis methods are also described in Section 3.2. Trend results are presented in Section 3.3 with respect to the catchment and climatic characteristics of the streamflow location. Finally, implications for studies of streamflow observations and global-scale modelling efforts are discussed in Section 3.4.

3.2 Data and methodology

3.2.1 Overview of streamflow database

High-quality time series of streamflow are critical for change detection studies (Sheng and Linghui, 2012). However as indicated in the introduction, due to data limitations, most flood hazard change-detection studies are

conducted at national and regional scales. These studies generally have relied on reference hydrometric network databases that have undergone extensive quality-checking procedures, and thus are less likely to be affected by low-quality time series (Whitfield et al., 2012). Furthermore, several of the datasets contain streamflow records only for catchments with minimum local anthropogenic influences (e.g. land use change, impoundments), so that any observed changes are more likely to be attributed by climatic causes (Burn et al., 2012). Examples of reference databases that have been widely used include the UK Benchmark Network (Hannaford and Marsh, 2006, Hannaford and Marsh, 2008, Stahl et al., 2010); the Unites States Hydro-Climatic Data Network (Douglas et al., 2000, Lins and Slack, 1999, Schilling and Libra, 2003, Mallakpour and Villarini, 2015); the Canadian Reference Hydrometric Basin Network (Yue et al., 2003, Yue and Wang, 2002); and the Australia Hydrologic Reference Stations (Franks and Kuczera, 2002, Micevski et al., 2006, Turner et al., 2012).

In contrast to the above datasets, no reference database exists for global streamflow data, due to the diversity of instrumentation, collection and archiving methods used across different countries, and limitations in documentation (Hannah et al., 2011). The main dataset used for global-scale streamflow investigations is the Global Runoff Data Centre (GRDC) database, which was initiated in 1988 by the World Meteorological Organisation. The GRDC is maintained at the German Federal Institute of Hydrology in Koblenz (http://grdc.bafg.de), and holds records of 9,213 stations across the globe (spatial coverage shown in Figure 3-1), with an average time series length of 42 years per station (GRDC, 2015). However, interpretation of any trends from the GRDC data requires caution since the catchments may or may not be affected by anthropogenic activities, and quality control processes are likely to vary depending on the country of origin.

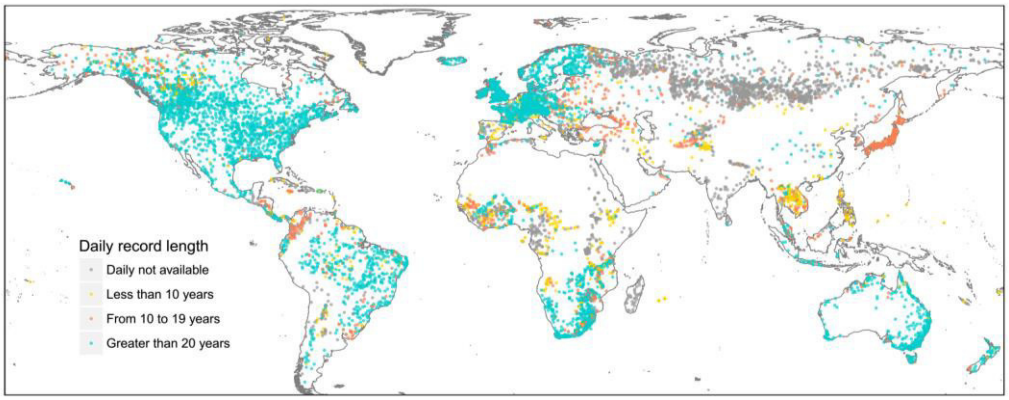


Figure 3-1: Global coverage and record length of GRDC streamflow stations

3.2.2 Reference periods used for analysis

To facilitate the detection of change in flood magnitude, the minimum data record length should be longer than decadal periodicity of hydrological cycles (Sheng and Linghui, 2012). Of the 9,213 records in the GRDC, the time series were initially filtered using the following procedure:

- For each station, 'flood' flows were obtained as the maximum daily value for each calendar year. In this step, the percentage of missing data was also estimated, and the annual maximum streamflow for that year was classified as 'missing' if more than 10% of days within that year were not available.
- Stations with less than 30 years of available annual maximum data (not necessarily continuous) were then removed (comprising 5,620 stations in total). In addition, stations with instances of two or more identical annual maximum streamflow values were also identified and manually checked through visual inspection, leading to the removal of a further 35 stations due to suspect quality. These gauges were mostly located in Africa and South America.

The remaining 3,558 stations represent the most comprehensive observation-based record of annual maximum streamflow at the global scale currently available. However, the number of annual maxima for these stations varies significantly through time (as shown in Figure 3-2), reaching its peak

in the late 1970s to 1990s followed by a decrease in the 2000s owing to data collation rather than discontinuation of gauges.

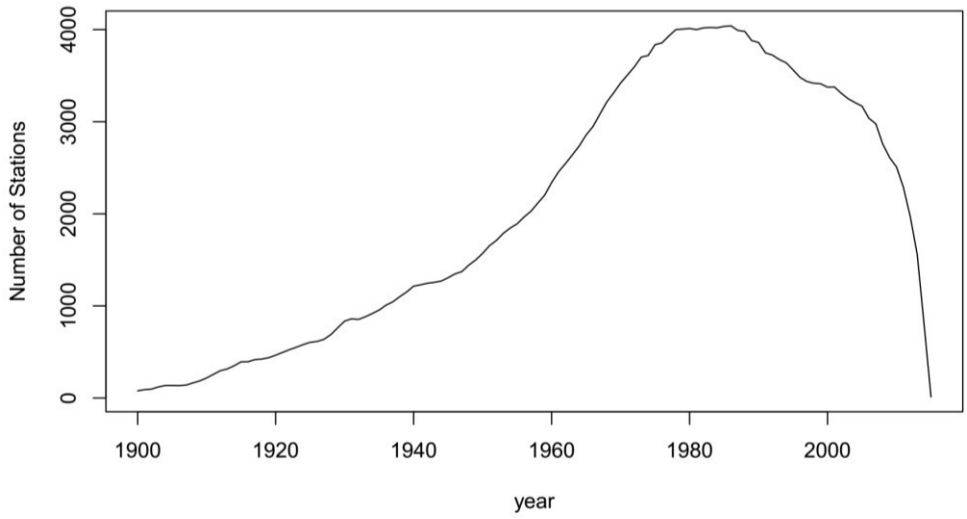


Figure 3-2: The total number of stations with at least 30 years of available data that have suitable annual maximum streamflow records in any given year

Given the changes in data availability with time, three separate reference periods were identified for this study:

- Dataset A1 (1,907 stations) comprises stations with at least 38 years annual maximum streamflow records over the 1966-2005 period (average record length of 39.7 years). This strict admission criterion (no more than 5% missing data) helps ensure the consistency in identifying changes in flood hazard over the reference period (while not compromising spatial coverage) with a good balance between the length of data series and the number of stations.
- Dataset A2 (3,478 stations) comprises stations with at least 30 years annual maximum streamflow over the 1955-2014 period (average record length of 47.6 years). This more generous admission criteria (up to 50% missing data is allowed) leads to a larger set of stations and a better spatial coverage. However, this dataset is also accompanied by a significant reduction in terms of data continuity.
- Dataset A3 (721 stations) comprises stations with at least 80 years annual maximum streamflow over the 1900-2014 period (average record length of 93.0 years). This long period together with fairly

strict admission criteria (less than 30% missing data allowed) constrains the available stations to a smaller geographic area (largely North America and Europe) but with a better temporal coverage.

Of these three periods, dataset A1 represents the best compromise among data length, completeness and availability of stations. Therefore, this dataset is used as the primary dataset for subsequent analyses in this paper, with datasets A2 and A3 analysed for comparative purposes in Section 3.3.1.

3.2.3 Filtering streamflow data based on catchment characteristics

One challenge in detecting trends in annual maximum streamflow is that streamflow measurements are sensitive to local anthropogenic influences (e.g. human regulation over the upstream catchment area). To reduce the influence of human-induced impact when evaluating trends, many studies have used catchment area as a selection criterion. Specifically, small catchments with areas less than 1000 km^2 have been preferred (Ishak et al., 2013, Stahl et al., 2010, Hannaford et al., 2013) as these catchments are considered less likely to have been extensively modified.

In this study, an alternative approach was taken whereby the upstream catchment area of each station was delineated and used to derive metadata such as the existence of large dams, forest-cover changes and climatological classification, which can be used to provide insight into the likelihood of anthropogenic changes in each catchment. To this end, catchment boundaries of the contributing upstream region for each gauge were obtained from the GRDC. A number of the GRDC streamflow records did not have catchments delineated, and for these catchments the HydroSHEDS flow direction data was used for the delineation (http://hydrosheds.cr.usgs.gov), which is available at 15 arc-seconds resolution (Lehner et al., 2006). The Viewfinder Panoramas elevation product (http://viewfinderpanoramas.org/) was also used as an alternative to address lack of coverage in the polar region and on some islands in the HydroSHEDS dataset. Having delineated each catchment, it is possible to associate catchment features with each streamflow gauge (see following sections for details).

3.2.3.1 The Global Reservoir and Dams (GRanD) database

The Global Reservoir and Dams (GRanD) database (Lehner et al., 2011) was used with delineated catchment boundaries to determine whether an upstream storage was present. The GRanD database contains storages with a capacity of at least 0.1km^2, representing 6,862 records of reservoirs and their associated dams. Although this is a valuable dataset for investigating the influence of dams on flood hazard, it is cautioned that the data is obtained from various research groups on a voluntary basis, and thus it is unlikely to provide a comprehensive description of all dams globally.

The catchment boundary positions of all 1,907 stations in dataset A1 were compared with the dam coordinates to detect the presence of dam(s) and classify stations into two sub-datasets:

- Dataset B1 (1,143 stations): Stations with no dams within the upstream boundary. This dataset is referred to as the "no dams" dataset.
- Dataset B2 (764 stations): Stations with at least one dam within the upstream boundary. This dataset is referred to as the "dams" dataset, which might be affected by dam construction or operation, leading to misinterpretation of detected changes.

3.2.3.2 Köppen-Weiger climate classification

World maps of the Köppen-Weiger climate classification (Rubel and Kottek, 2010) were used to identify the climate classification for each catchment. This dataset was developed using a gridded monthly temperature product provided by the Climatic Research Unit (CRU) of the University of East Anglia (Mitchell and Jones, 2005) and a gridded precipitation product (GPCC's full data reanalysis version 4) obtained from the Global Precipitation Climatology Centre (GPCC) for 1901–2007 (Fuchs et al., 2007). These observation-based datasets cover the global land areas excluding Greenland and Antarctica and have undergone a series of quality control steps to avoid inhomogeneities in the gridded data. The world map of Köppen-Weiger climate classification comprises 31 climate classes, which were

adjusted into nine sub-climatic groupings and used in this study. The spatial distribution of the nine sub-climatic classifications is illustrated in Figure 3-3.

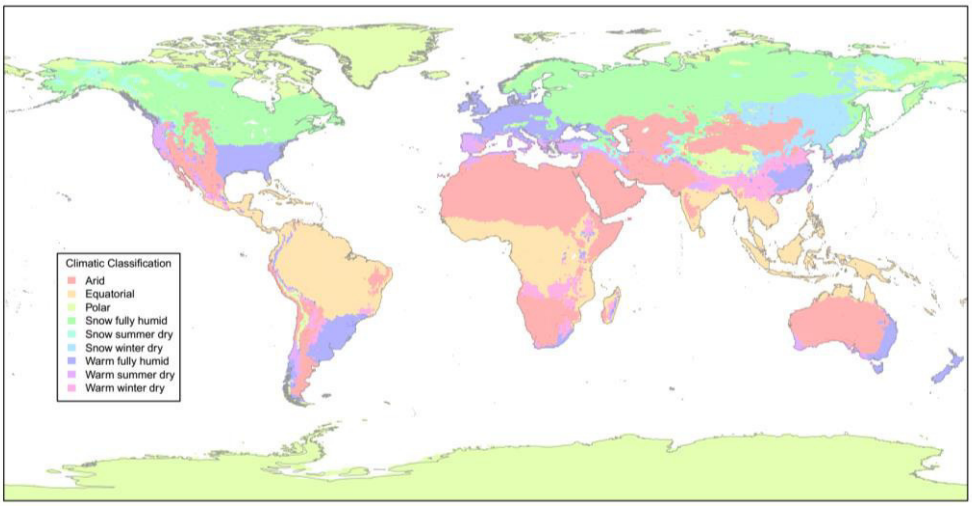


Figure 3-3: Spatial distribution of nine groupings of Köppen climate types used in this study based on Rubel and Kottek (2010).

The nine sub-climatic groupings were initially developed from the five major climatic groups. These major groups comprise equatorial climates (A), arid climates (B), warm temperate climates (C), snow climates (D) and polar climates (E). Across these five major groups, temperate climates (C) and snow climates (D) were regions with the highest density of available runoff stations, so that each of these classifications were further subdivided into three climate groups to create the nine-group classification system. The main features of the classified climates are provided in Table 3-2.

Table 3-2: Main features of nine climatic groups, based on the Köppen-Weiger climate classification system.

Climate groups	**Description**
Equatorial climates (A)	Tmin >= + 18oC
Arid climates (B)	Pann < 10 Pth
Warm temperate climates (C): Lowest monthly mean temperature between -3 oC and 18oC	
Warm summer dry (Cs)	Psmin < Pwmin, Pwmax > 3Psmin and Psmin < 40 mm
Warm winter dry (Cw)	Pwmin < Psmin and Psmax > 10 Pwmin
Warm fully humid (Cf)	neither Cs nor Cw
Snow climates (D):	Lowest monthly mean temperature <= -3 oC
Snow summer dry (Ds)	Psmin < Pwmin, Pwmax > 3Psmin and Psmin < 40 mm
Snow winter dry (Dw)	Pwmin < Psmin and Psmax > 10 Pwmin
Snow fully humid (Df)	neither Ds nor Dw
Polar climates (E):	Tmax <= 10 oC

T_{min} (°C) : Lowest monthly mean air temperature

T_{max} (°C) : Highest monthly mean temperature

P_{ann} (mm/year) : Accumulated annual precipitation

P_{wmin} (mm/month) : Lowest monthly precipitation for the winter half-year

P_{smin} (mm/month) : Lowest monthly precipitation for the summer half-year

P_{wmax} (mm/month) : Highest monthly precipitation for the winter half-year

P_{smax} (mm/month) : Highest monthly precipitation for the summer half-year

P_{th} (mm) : Dryness threshold

where:

$$P_{th} = \begin{cases} 2*T_{ann} & \text{if at least 2/3 of annual precipitation occurs in winter,} \\ 2*T_{ann}+28 & \text{if at least 2/3 of the annual precipitation occurs in summer,} \\ 2*T_{ann}+14 & \text{otherwise.} \end{cases}$$

3.2.3.3 The Global Forest Change dataset 2000–2012

The Global Forest Change dataset 2000–2012 version 1.0 (http://earthenginepartners.appspot.com) was used to derive the proportions of change in forest area, in which "forest" is defined for locations with trees at least 5m tall (Hansen et al., 2013). This satellite-derived product was obtained from Landsat 7 ETM+ images based on the Normalized Difference Vegetation Index (NDVI), and is available at one arc-second resolution. The Global Forest Change dataset comprises several products, and the following two layers were used in this study:

- Global forest cover loss 2000–2012 (loss): Forest loss during the period 2000–2012 was identified when a disturbance (or a change from a forest to a non-forest state) was detected. Each cell in this dataset was assigned either a value of 1 (loss) or 0 (no loss).
- Global forest cover gain 2000–2012 (gain): Forest gain during the period 2000–2012 was identified when the inverse of loss (or a change from non-forest to forest state) was detected. Each cell in this dataset was assigned either 1 (gain) or 0 (no gain) value.

To assess forest loss or gain, the catchment boundary associated with each stream gauge was used to extract the number of one arc-second pixels that experienced loss (f_l) and gain (f_g) in forest cover over 2000–2012. The percentage of each catchment area that has undergone forest change (f_c) was then derived from the extracted values of each catchment as shown in Equation (1).

$$f_c = \frac{f_g - f_l}{A} * 100\ (\%) \qquad (1)$$

where A is the total number of one arc-second size cells of the catchment.

Resulting from this calculation, f_c can be positive (indicating that the catchment experienced an expansion in forest cover area during 2000 – 2012), negative (indicating that the catchment experienced a reduction in forest cover area during 2000 – 2012) or zero (indicating that no change has occurred during 2000 – 2012). The R-package gfcanalysis (Zvoleff, 2015) was used to download and analyse this data.

3.2.4 Statistical techniques to assess changes to annual maximum streamflow records

To test for monotonic changes in the annual maximum streamflow time series, a Mann-Kendall test was applied separately to each location (Wilks, 2011). In previous studies (as summarised in Table 3-1), statistically significant trends were typically reported at the 10% two-sided significance level (with the exception of those studies conducted over the United States). To compare this study with existing studies, the null hypothesis is rejected if the two-sided *p*-value of the test statistic (Kendall's τ) is lower than 0.1. Depending on the value of τ (i.e. positive or negative), we then infer that there is a monotonic increasing/decreasing trend in the annual maximum streamflow over time.

To assess whether the proportion of trends of hydrology time series was significant over a specific group of observations, a field significance resampling procedure (Wilks, 2011) was applied (Westra et al., 2013, Ishak et al., 2013, Kiktev et al., 2003, Von Storch and Navarra). The general steps of this approach are summarised as:

(1) The reference period, e.g. [1966, 1967, 1968, 1969… 2005], was resampled with replacement to create a new set with the same length but different year-order from the original period, for example [2005, 1966, 2001, 1998… 1966].

(2) The observations at each site were rearranged corresponding to the resampled set of years obtained from step (1) to obtain a new sample of annual maximum streamflow, where any temporal structure was removed but the spatial dependence was preserved.

(3) The Mann-Kendall test was applied to the resampled streamflow at the 10% two-sided significance level, and the percentage of stations showing significant (both increasing and decreasing) trends was recorded.

(4) Steps (1) to (3) were repeated 1000 times to obtain the percentage distribution of stations showing significant trends, representing the distribution of the null hypothesis.

The above approach is appropriate when the data are serially independent; however in this case stations exhibited mild levels of autocorrelation, with a global median value of 0.073, 0.068, and 0.078 for datasets A1, A2 and A3. To account for the effect of autocorrelation on the field significance results across different groups of stations, a moving-blocks bootstrap approach was adopted (Kiktev et al., 2007, Alexander et al., 2006). The distinction of the moving-blocks bootstrap in comparison to the conventional bootstrap lies in the first step of the resampling procedure (step 1, above), when consecutive years of length L are resampled in order to build up a synthetic sample with a similar level of autocorrelation to the original data. As autocorrelation in the time series varies across different sites, the block length is recommended to be determined based on individual time series (Politis, 2003), and thus L also varies across different sites. To identify optimal block length L for individual annual maximum streamflow time series, the automatic block-length selection procedure (Politis and White, 2004) was implemented using the R script available at http://www.math.ucsd.edu/~politis/SOFT/PPW/ppw.R).

The moving-blocks bootstrap resampling procedure that incorporated the automatic block-length selection process was used for field significance analysis across the experiments in this study and is referred to as the "moving-blocks bootstrap procedure" throughout this paper. Percentages calculated from the observed datasets were compared to resampled distributions generated from the moving-blocks bootstrap procedure to determine whether the annual maximum data of a specific group of stations exhibited monotonic trends. The 10% two-sided significance level of the Mann-Kendall test implies that about 5% of stations would show significant increasing/decreasing trends by random chance. The null hypothesis of no

change is rejected when the observed percentage lies outside the 90% confidence interval of the resampled distributions.

Finally, an alternative method to assess the role of catchment characteristics on the presence or absence of trends is to assess whether various attributes of catchments with statistically significant increasing and decreasing trends are similar or not. To do this, a *t*-test at the 5% significance level was used. If the *p*-value of a specific test is lower than 0.05, the null hypothesis is rejected and there is evidence to support the statement that there is a significant difference in the attribute subjected to the *t* test between stations showing increasing trends and stations showing decreasing trends.

3.2.5 Overall study approach

Based on the above datasets and statistical testing approaches, four main experiments were developed to detect changes in annual maximum streamflow over different spatial domains, and to link any detected changes with climatic and catchment characteristics. The purpose of each experiment, the required datasets and the statistical tests used, with each experiment summarised as follow:

- **Identify the global patterns of change in annual maximum streamflow and the influence of reference periods.** Three datasets (A1, A2, and A3) were used to assess changes in the annual maximum flow series and the response of detected trends to changing reference periods. In this experiment, the Mann-Kendall test was applied to each location and the percentage of stations showing significant trends was recorded for each dataset. These numbers were analysed to identify the importance of choosing a reference period in analysing trends of floods at the global scale.
- **Identify the importance of dams to changes in annual maximum streamflow.** In this experiment, the A1 dataset was divided into two subsets (B1 “no dams” and B2 “dams”). The Mann-Kendall test was applied at each location and a bootstrap procedure was applied to create the null hypothesis distribution for each dataset. The comparison between observed percentages of stations showing

significant trends and the null hypothesis distribution (obtained from the moving-blocks bootstrap approach) across the three different datasets was used to assess whether the presence of dams has a significant influence on the results of the analysis.

- **Identify the relationship between geographical or climatic characteristics and detected trends.** Dataset B1 ("no dams") was analysed using the Mann-Kendall test and moving-blocks bootstrap procedure. To contrast the relationship between geographical or climatic characteristics and the detected trend, no dam stations were divided into six continents (Africa, Asia, Australia (and some islands), Europe, North America, South America) or nine climatic regions (arid climate, equatorial climate, polar climate, warm summer dry climate, warm winter dry climate, warm fully humid, snow summer dry climate, snow winter dry climate, snow fully humid). Each group was subjected separately to the same statistical analysis to identify differences in significant changes of annual maximum streamflow records across different continents and climatic groups.

- **Assess the effect of catchment area and forest cover changes on trends in annual maximum streamflow.** These two characteristics were chosen as they can affect the amount of water available for runoff. To assess changes to the relationship between flood and catchment characteristics, two separate experiments were conducted for each dataset:

  (1) **Field significance analysis for stations with different catchment characteristics.** In this experiment, all stations were divided into five groups with similar forest cover change/catchment sizes using the 20, 40, 60 and 80 quantiles of the frequency distribution. Field significance analysis was applied to these groups to identify whether a group exhibits a significant percentage of stations showing significant trends.

  (2) **Apply a *t* test for differences between stations showing increasing and decreasing trends.** Here all stations showing

significant trends were divided into two groups: a group of stations with increasing trends and a group with decreasing trends. A *t* test at the 5% significance level was conducted to assess whether the catchment characteristics of increasing trend and decreasing trend groups were indistinguishable. If the *p*-value is lower than 0.05, the null hypothesis is rejected and there is evidence to support that the catchment characteristics between these two groups are different. Based on the distribution of catchment characteristics for these two groups, relationships of these variables to changes in annual maximum streamflow records can be identified.

Table 3-3 Experiments conducted, and their associated datasets.

Experiment	Datasets required	Statistical tests	Results
Global pattern of changes in annual maximum streamflow	- Three different streamflow datasets: A1, A2, A3 (see section 3.2.3.1)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure)	Section 3.3.1
Impacts of dam to changes in annual maximum streamflow	- GRanD database was used to classify A1 dataset into B1 (no-dam stations) and B2 (dam stations) datasets - B1 (no dam stations) and B2 (dam stations)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure)	Section 3.3.2
Relationships between changes in annual maximum streamflow and geographical/climatic characteristics	- Köppen-Weiger climate classification system was used to assign stations into different climatic groups - Dataset B1 (no dam stations)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure)	Section 3.3.3
Relationships between changes in annual maximum streamflow and catchment characteristics	- Catchment size and proportion of forest cover changes (2000-2014) were used to assign stations into different catchment characteristics group - Dataset B1 (no dam stations)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure) - *t* test	Section 3.3.4 and Section 3.3.5

3.3 Results

3.3.1 Patterns of change to global annual maximum streamflow and the influence of different reference periods

We start by investigating the number of stations with statistically significant increasing and decreasing trends at the 10% (two-sided) significance level, and compare to what would be expected under the null hypothesis (obtained from moving-blocks bootstrap procedure). The percentage of stations with significant trends is presented in Figure 3-4 for each of the three reference periods, with the null hypothesis distribution generated from 1,000 conventional bootstrap iterations given as a histogram. As can be seen, over the main reference period (dataset A1; 1966 – 2005), there were 7.1% of stations with statistically significant increasing trends (corresponding to 136 stations; Figure 3-4a), and 11.9% of stations with statistically significant decreasing trends (corresponding to 226 stations; Figure 3-4b). The percentage of stations exhibiting statistically significant increasing trends is consistent with the null hypothesis of no change on average across the global dataset, whereas the percentage of stations showing significant decreasing trends is inconsistent with the null hypothesis.

Comparing the results from the main reference period (dataset A1) to the other two reference periods, it is evident that the percentage of stations with significant increasing trends changes from 7.1% (Figure 3-4a), to 11.8% for dataset A2 (Figure 3-4c) and 16.4% for dataset A3 (Figure 3-4e). A similar pattern is evident for stations with decreasing trends, with the percentage of stations showing decreasing trends changing from 11.9% (Figure 3-4b), to 15.4% for dataset A2 (Figure 3-4d) and 22.9% for dataset A3 (Figure 3-4f). With the exception of increasing trends in dataset A1, all results are field significant at the 10% (two-sided) significance level, even when the moving-blocks bootstrap (block size of two) procedure was adopted. Interestingly, more stations showed statistically significant decreasing trends than statistically significant increasing trends for all datasets.

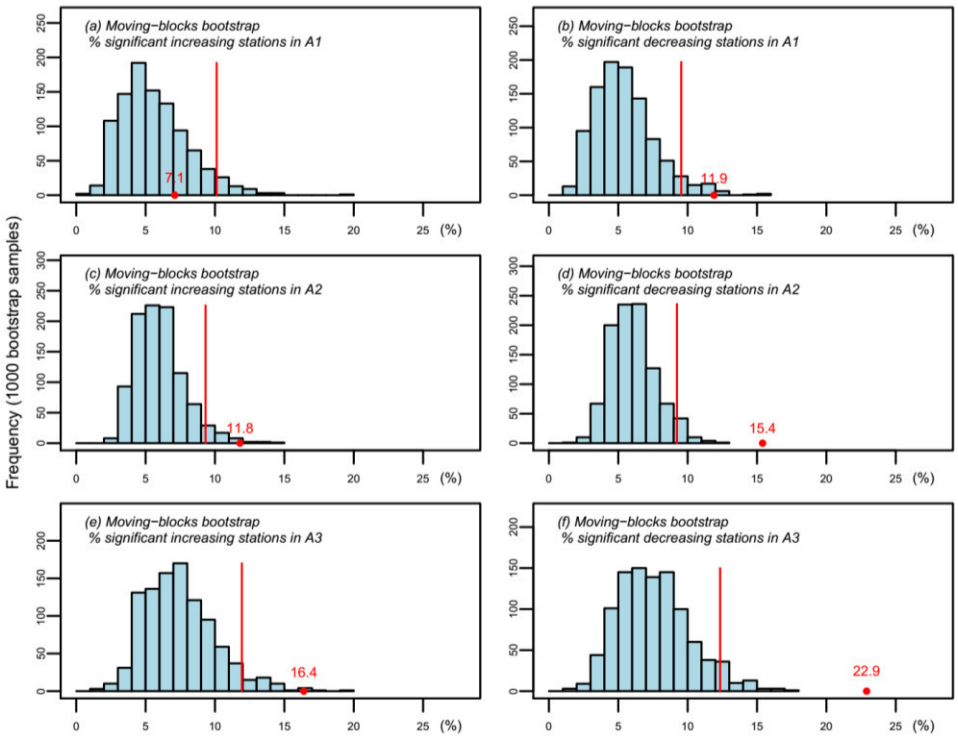


Figure 3-4: Percentage of stations showing significant trends based on the Mann–Kendall test with each row displaying increasing and decreasing trends respectively. For dataset A1 (Figure 3-4a and 4b); dataset A2 (Figure 3-4c and 4d); and dataset A3 (Figure 3-4e and 3-4f)

One possible explanation for the different results between datasets is the effect of different data lengths on the statistical power of the Mann-Kendall tests, with datasets A1, A2 and A3 having an average of 39.7, 47.6 and 93.0 years of data, respectively (see Section 3.2.2). For example, of the 652 stations that were common to datasets A1 and A3, there were 105 stations (16.1% of the sample) exhibiting significant changes (either increases or decreases) for dataset A1 compared to 257 stations (39.4% of the sample) exhibiting significant changes for dataset A3, indicating a greater likelihood of trend detection for the longer dataset. However, the composition of the stations that were statistically significant was not consistent across the group: in particular, considering again those stations that were common between datasets A1 and A3, approximately half of the stations that exhibited significant trends in group A1 were no longer significant in group A3. Similar patterns were found when comparing common stations between datasets A1 and A2, confirming the potentially strong role of sampling variability combined with potential systematic changes over time at individual gauges. This result is consistent with other studies that find a strong influence of the

reference period window on the results of trend detection analyses (Hall et al., 2014, Merz et al., 2012a, Kundzewicz and Mondiale, 2000, Hannaford et al., 2013, Lins and Slack, 1999).

An alternative explanation of the differences between datasets is the potential for shifts in the spatial distribution of the trend results. Spatial plots of stations showing statistically significant increasing and decreasing trends are presented in Figure 3-5. Starting with dataset A1 (upper panel of Figure 3-5), some regional clustering of the trending stations was evident, with a large number of stations with significant decreasing trends in North America (particularly to the west of the continent), and regions with increasing trends concentrated in parts of Europe (e.g. the UK, France and Germany) and parts of eastern North America.

When these results are compared to the other datasets (middle and lower panels of Figure 3-5), it is apparent that although data coverage varies significantly between the datasets, there are important consistencies in the spatial patterns of the trends. For example, the western part of North America and the data-covered regions of Australia (particularly the east coast of the continent) showed a number of stations with decreasing trends, regardless of the dataset used. Similarly, regions with increasing trends—for example in Europe, eastern North America, parts of South America and southern Africa—were generally consistent across the datasets (although data from South America and southern Africa were not available for dataset A3). With a better spatial coverage, dataset A2 also revealed information for regions not adequately covered by datasets A1 and A3: in particular, the number of stations showing increasing trends was outnumbered by the stations showing decreasing trends over Asia with the exception of the far eastern region of Russia, and all stations located in western Africa showed significant decreasing trends. However, due to data limitations over most countries in these regions, and the generous data admission criteria for dataset A2 (allowing up to 50% of missing data), these results should be interpreted with care.

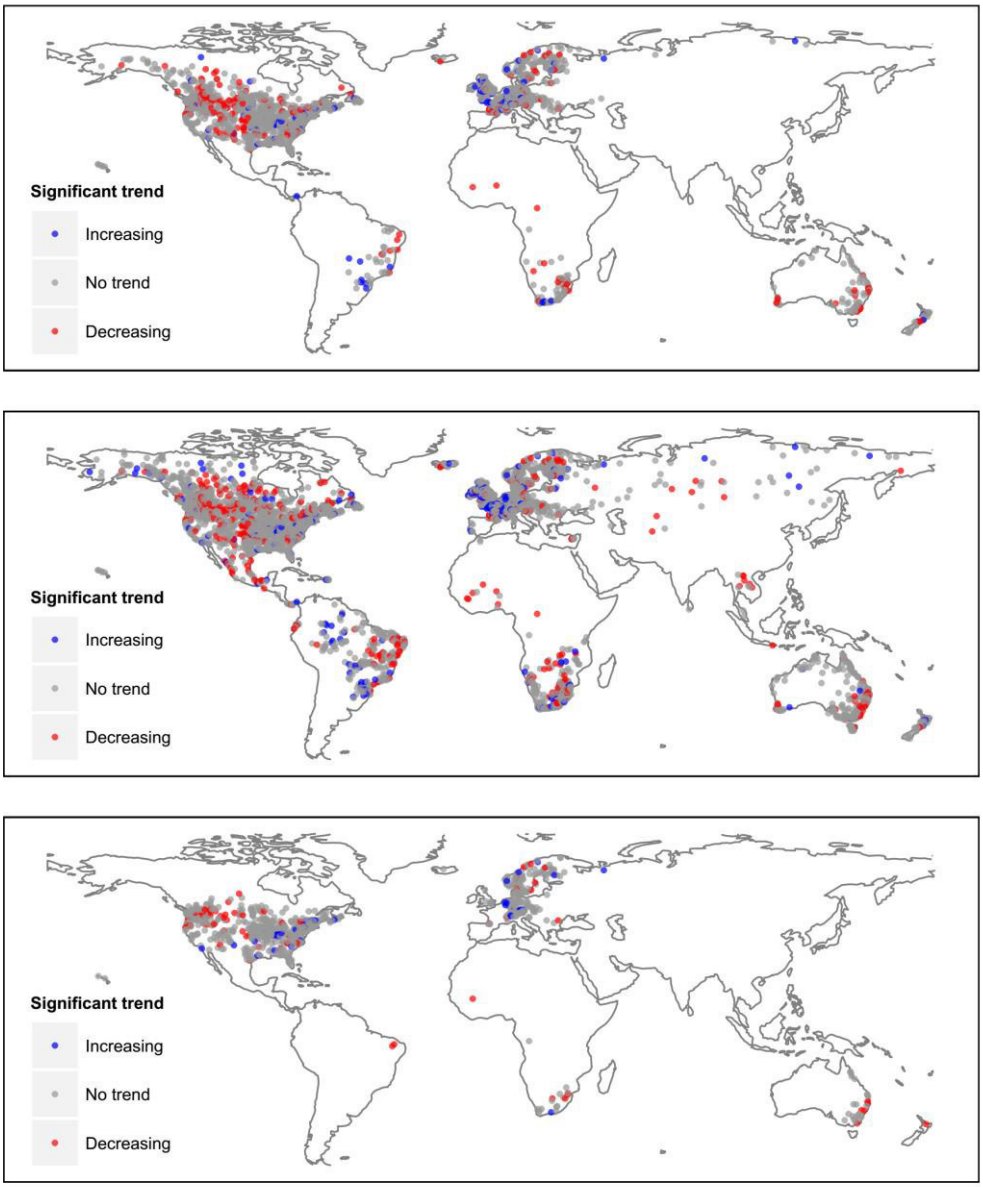


Figure 3-5: Results for trends in magnitude of floods events for dataset A1 (upper panel), A2 (middle panel) and A3 (lower panel). Blue (red) filled dots indicate stations with statistically significant increasing (decreasing) trends at the 10% level. Grey dots indicate time series that did not exhibit statistically significant changes at the 10% level.

Based on the above analyses, it can be concluded that although individual stations exhibiting statistically significant trends were often different between reference periods, the above spatial analysis indicates some consistency between the reference periods at the regional scale. Furthermore, stations showing significant increasing trends were outnumbered by stations showing significant decreasing trends in all three datasets, indicating that although there may be evidence of regional increasing trends in flood hazard, the

hypothesis that there is a significant increase in flood hazard when averaged over all the data-covered regions of the globe is not supported by this analysis.

There are multiple systematic factors that could explain these findings, which are investigated in the following sections, such as the influence of dams and other forms of human activity. As dataset A1 provides a reasonable compromise between data length and spatial coverage, we focus on this dataset in the following sections.

3.3.2 Impact of dams on the number of stations showing significant changes and the importance of resampling technique

To assess impacts of dams on trends in annual maximum streamflow records, field significance analyses using moving-blocks bootstrap were applied separately to the "no dams" group (B1) and "dams" group (B2), and the results compared with the full reference period (A1). The results are given in Figure 3-6, and show that 46 stations (6.0%) had statistically increasing trends in the "dams" group compared to 90 stations (7.9%) for the "no dams" group, whereas 105 stations (13.7%) had statistically significant decreasing trends for the "dams" group compared to 121 stations (10.6%) for the "no dams" group. The percentage of stations showing statistically significant increasing trends was not field significant (Figure 3-6a and 3-6e), whereas the percentage of stations showing statistically significant decreasing trends was field significant (Figure 3-6c and 3-6g) when either B1 or B2 was used. The results were therefore generally consistent with the reference dataset A1, indicating that the effect of dams does not appear to have a substantive influence on the overall trend results.

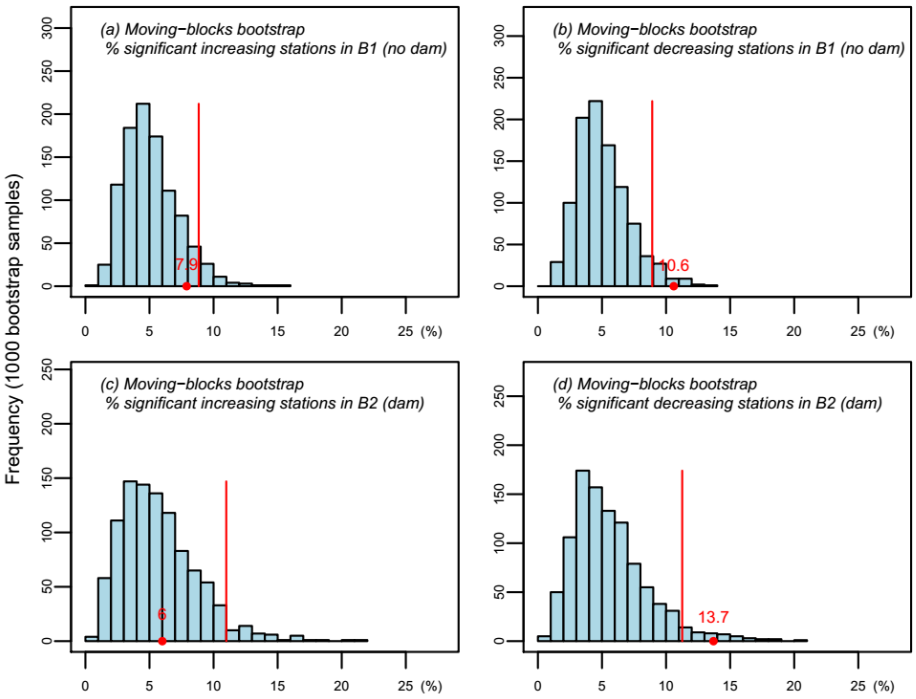


Figure 3-6: Percentage of stations showing significant trends based on the Mann–Kendall test for 1,143 stations in "no dams" (B1) dataset (Figure 3-6a and 6b), and 746 stations in "dams" (B2) dataset (Figure 3-6c and 3-6d). Left panels show the results for percentage of stations showing significant increasing trend while right panels show results for percentage of stations showing significant decreasing trend. The histogram represents the distribution of percentages obtained from 1000 moving-blocks bootstrap iterations; the red dot indicates the observed value while the red line indicates the 95th percentile.

Although the above results suggest that the presence of large dams does not substantially influence the overall trend results, from a hydrological perspective, large dams would be expected to have a significant effect on flood flows, as in many cases the dams are designed to reduce flood magnitude and the flood damage on human assets (FitzHugh and Vogel, 2011, Lajoie et al., 2007, Jaramillo and Destouni, 2015). On this principle, remaining analyses are focused on the "no dams" case (i.e. dataset B2) to minimise the influence of large hydraulic structures on any trend results.

3.3.3 Relationships between trends in floods and geographical/climatic characteristics

This section evaluates whether trends can be associated with the geographical or climatic characteristics of stations. To do this, all stations were divided into either six continents (representing similar geographical characteristics) or nine Köppen climate zones (representing similar climatic

characteristics). The results of the continent-based analysis are shown in Table 3-4, with three continents (North America, Australia and Africa) showing more stations with decreasing trends than increasing trends, and three continents (South America, Asia and Europe) showing more stations with increasing trends than decreasing trends. However, as with the results shown in Figure 3-5, the data were not evenly distributed between continents. For example, the whole of Asia is covered by only two streamflow locations, and thus is not sufficient to constitute a representative sample.

Table 3-4: Summary of Mann-Kendall test and field significance analysis across different continents.

	Asia	**North America**	**Europe**	**South America**	**Australia**	**Africa**	**Global**
Available stations	2	418	560	26	79	58	1,143
Number of increasing trend stations	1 (50.0%)	11 (2.6%)	67 (* 12.0%)	6 (* 23.1%)	2 (2.5%)	3 (5.2%)	90 (7.9%)
Number of decreasing trend stations	0 (0%)	55 (* 13.2%)	37 (6.6%)	5 (* 19.2%)	15 (* 19.0%)	9 (15.5%)	121 (* 10.6%)

(: field significant at 10% level)*

The results of the field significance analysis using the moving-blocks bootstrap technique for each continent (which accounts for the number of stations in the record) are consistent with what was found at the global scale (Section 3.3.1), with the percentages of stations showing significant decreasing trends being field significant in three continents (13.2%, 19.2% and 19.0% for North America, South America and Australia respectively), whereas Europe and South America show statistical evidence of increasing trends (12.0% and 23.1% respectively).

Focusing on continents well-represented in the database (i.e. Europe and North America), the most notable trends are in the western part of North America (as shown in Figure 3-7), where there are approximately ten times as many stations with significant decreasing trends compared to significant increasing trends. This was also shown in previous studies of the United States (Lins and Slack, 1999), and Canada (Zhang et al., 2001, Whitfield, 2001, Westmacott and Burn, 1997, Cunderlik and Ouarda, 2009), which generally have found a decrease in the magnitude of floods over time that might be at least partially attributable to changes in snowmelt timing. In other

parts of North America and Europe, stations with significant increasing or decreasing trends appear to be distributed more randomly and this is consistent with previous studies of the United States (Lins and Slack, 1999, Villarini et al., 2009, Mallakpour and Villarini, 2015) and Europe (Hannaford et al., 2013, Stahl et al., 2011).

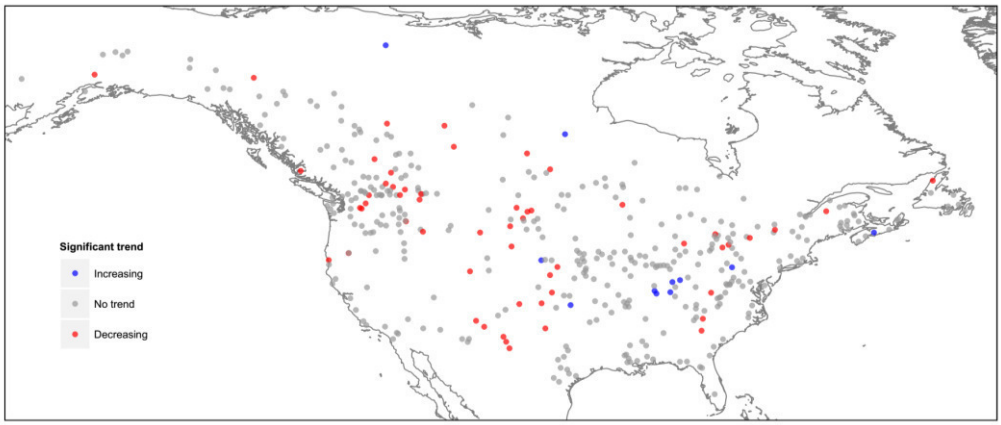


Figure 3-7: Trends in magnitude of annual maximum flow events for stations in the B1 dataset over North America. Blue (red) filled dots indicate stations with a statistically significant increasing (decreasing) trend at the 10% level. Grey dots indicate that the time series that did not exhibit statistically significant changes at the 10% level.

In addition to grouping the results by continent, stations are also grouped by climatic zone (Table 3-5), and the results provide further consistent evidence to what have been found at global scale. As can be seen, field significant decreasing trends are observed with the moving-blocks bootstrap technique in: equatorial climates, arid climates, and snow fully humid climates. In contrast, warm fully humid is the only climate zone shows field significant increasing trends.

Table 3-5: Percentage of stations showing significant increasing (decreasing) trend at the 10% significance level over climatic regions

Climate groups	No. of stations	Number of increasing trend stations	Number of decreasing trend stations
Equatorial climates (A)	24	3 (12.5%)	5 (* 20.8%)
Arid climates (B)	52	2 (3.8%)	19 (* 36.5%)
Warm summer dry (Cs)	50	1 (2.0%)	7 (14.0%)
Warm winter dry (Cw)	22	0 (0%)	3 (13.6%)
Warm fully humid (Cf)	562	60 (* 10.7%)	38 (6.8%)
Snow summer dry (Ds)	20	0 (0%)	2 (10.0%)
Snow winter dry (Dw)	1	0 (0%)	0 (0%)
Snow fully humid (Df)	386	20 (5.2%)	45 (* 11.7%)
Polar climates (E)	26	4 (15.4%)	2 (7.7%)
Total	1,143	90 (7.9%)	121 (* 10.6%)

(: field significant at 10% level)*

Table 3-5 also demonstrates the inhomogeneity in data coverage across climatic regions. In particular, more than 83% of stations are located in the climatic regions "warm fully humid" and "snow fully humid", corresponding to the large proportion of these regions in North America and Europe.

3.3.4 Relationships between trends in floods and catchment area

Catchment area has been used in a number of studies as an indicator of the magnitude of land-use change (Ishak et al., 2013, Stahl et al., 2010, Hannaford et al., 2013), and we now explore the influence of area on the overall trend results. To identify the relationship between trends in floods and catchment area, the 1,143 "no dam" stations were divided into five groups with an equal number of samples based on their catchment area. Similar to previous sections, field significance was assessed using moving-blocks bootstrap procedure. Of all 1,143 stations, catchment area varies from 1km^2 to 500,000 km^2, with more than 60% of the catchments having areas less than ~1,400km^2. The 20, 40, 60 and 80 quantiles of the catchment area distribution were used as thresholds to create five different groups of stations based on their area: "very small", "small", "medium", "large" and "very large" catchments. As shown in Table 3-6, the percentage of stations with decreasing trends was found to be significant for the three latter groups, while the percentage of stations exhibiting statistically significant increasing trends was field significant for only "very small" group.

Table 3-6: Percentage of stations showing significant increasing (decreasing) trends at the 10% significance level over different catchment-size groups

	Catchment size groups				
	Very small catchments	**Small catchments**	**Medium catchments**	**Large catchments**	**Very large catchments**
Number of stations	229	228	229	228	229
Significant increasing	26 (*11.4%)	22 (9.6%)	21 (9.2%)	10 (4.4%)	11 (4.8%)
Significant decreasing	22 (9.6%)	11 (4.8%)	28 (*12.2%)	27 (*11.8%)	33 (*14.4%)
Size of catchment (km^2)	1 – 134	134-389	389-1,368	1,368-4,008	4,008-500,000

(: field significant at 10% level)*

A notable pattern in the percentages of stations showing significant trends over these catchment-size groups is that groups with larger catchment areas generally have stations with more (less) significant decreasing (increasing) trends. As illustrated in Figure 3-8, the percentage of stations showing

significant decreasing trends is three times the percentage of stations showing significant increasing trends in the last two groups (catchments with area above 1,368km^2). Even in stations in the "medium catchment" group, stations showing significant decreasing trends also outnumbered those showing significant increasing trends. In contrast, the "very small" and "small" catchment groups generally have more stations showing significant increasing trends than stations showing decreasing trends. An implication of this pattern is that the strategy of choosing stations using catchment area may have a substantial influence on the detected changes in annual maximum streamflow records. For example, analysing changes in annual maximum streamflow records using only small catchments (e.g. catchments with area less than 500km^2) will potentially lead to results where there are more stations showing significant increasing trends.

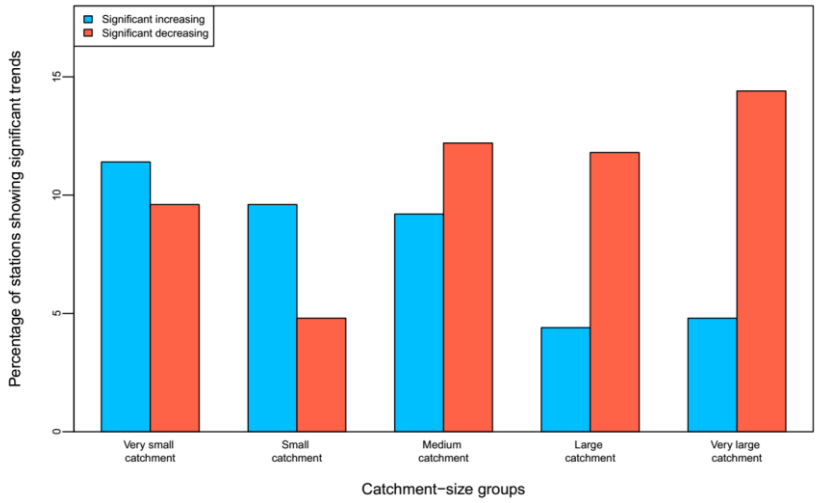


Figure 3-8: Bar chart of the differences in area between stations showing significant increasing and decreasing trends across five different catchment size groups.

As an alternative approach to analysing the impact of catchment size on trend results, all stations showing significant trends were divided into stations showing significant increasing trends and stations showing significant decreasing trends, to determine if the catchment areas between these two groups are different from each other. Interestingly, the results of the *t* test for catchments did not provide evidence to reject the hypothesis that there is no significant difference in catchment area between the two groups. Although the mean of catchment size of the significant decreasing trend group is higher

(9,932km^2 in comparison with 4,868km^2 in the significant increasing trend group), the null hypothesis is not rejected (p-value: 0.32). This indicates that although there is a clear pattern in trends as indicated by Figure 3-8, a high level of random variability is also present and the specific nature of the analyses is likely to have a large effect on the interpretation of the results—for example, studies focused on large basins.

3.3.5 Relationships between trends in floods and changes in forest cover rate

Finally, we investigate the role of forest cover on any identified trends in annual maximum streamflow, focusing again on the 1,143 "no dam" catchments. Each station was assigned into one among three different groups based on the percentage of changes in catchment area over 2000-2012: forest gain (f_c>0), no change (f_c = 0) and forest loss (f_c < 0), with results shown in Table 7. Interestingly, 68% of the catchments have experienced an expansion in forest cover, whereas only 24% of catchments have experienced a decline in forest cover, with an average proportion of total forest area cover change across all catchments being 1.1% over the 2000-2012 period. These numbers are slightly different from changes in forest area globally as reported in the recent Global Forest Resources Assessment (Keenan et al., 2015), which showed a declining proportion of 1.4% of global forest area over the same period. This may be caused by the inhomogeneous spatial coverage of the GRDC data, with the Keenan et al. (2015) analysis conducted on all continental areas except for Antarctica, whereas our results are based only on those regions with streamflow data.

To assess whether the number of stations showing significant changes is field significant, the moving-blocks bootstrap procedure was applied to each group separately. Firstly, considering the stations with increasing trends, the "forest loss" and "forest gain" datasets show a similar magnitude of increasing trends compared to the global scale "no dam" analysis (Section 3.3.2). In contrast, the dataset with "no change" in forest cover showed a higher percentage change, but this result is not statistically significant and may be due to the relatively small sample size of this dataset. Considering next the decreasing trends, the percentage changes were consistent with the

global scale "no dam" analysis for both "forest loss" and "forest gain" as field significance was detected.

As an alternative approach to testing the role of forest cover changes, the data were divided into quintiles by forest cover change (i.e. dividing the data at the 20, 40, 60 and 80 percentiles), and the field significance of these results were evaluated. The results (not shown) were consistent with the results presented in Table 3-7.

Table 3-7: Percentage of stations showing significant increasing (decreasing) trend at the 10% significance level over forest cover change groups

	Forest changes groups		
	Forest loss (f_c < 0)	**No change (f_c = 0)**	**Forest gain (f_c > 0)**
Number of stations	277 (24%)	89 (8%)	777 (68%)
Stations showing increasing trends	20 (7.2%)	10 (11.2%)	60 (7.7%)
Stations showing decreasing trends	32 (* 11.6%)	9 (10.1%)	80 (* 10.3%)

(*: field significant at 10% level)

3.4 Discussion and conclusions

Whilst acknowledging data concerns regarding data record length, spatial coverage and quality, this study provides an indication of changes to annual maximum streamflow in the data-covered regions of the globe. Three reference datasets representing different compromises between data length and spatial coverage were examined, and it was found that the choice of dataset had a significant influence in terms of whether individual stations had significant increasing or decreasing trends, indicating a large role of sampling variability and other factors on the station-level trend results. However, when comparing trend results of the three datasets at the global scale, it was found that a greater number of stations exhibited statistically significant trends (both increasing and decreasing) as the average length of record increased, with almost 40% of stations exhibiting statistically significant trends at the 10% significance level for the longest dataset (A3). This is at least partially explained by the increased statistical power of trend detection methods as the record becomes longer.

Although there were substantial differences in station-level results depending on the dataset used, the results were more consistent when viewed

at the regional scale. In particular, decreasing trends were observed for a large number of stations in western North America and the data-covered regions of Australia, and increasing trends in parts of Europe, eastern North America, parts of South America and southern Africa. These changes were also consistent with previous regional studies; for instance, previous studies have found that decreasing trends in the western part of North America (Zhang et al., 2001, Whitfield, 2001, Westmacott and Burn, 1997, Cunderlik and Ouarda, 2009, Lins and Slack, 1999) and in Australia (Ishak et al., 2013), while the eastern part of North America and Europe generally have a mixed pattern of changes (Lins and Slack, 1999, Villarini et al., 2009, Mallakpour and Villarini, 2015, Hannaford et al., 2013, Stahl et al., 2011). In addition to dividing the result by geographical location, consistent results were also found when the data were broken into separate climate groups, with decreasing trends particularly apparent in stations with equatorial, arid and snow fully humid climates (field significant at the 10% significance level) while increasing trends are present in only stations with warm fully humid climates.

Neither the presence of dams nor changes in forest cover had a large effect on the trend results (although most analyses were conducted using the "no dam" dataset to minimise the potential influence of large hydraulic structures on results). The lack of a conclusive finding regarding the influence of forest cover suggests that forest cover may not play a dominant role explaining trends in flood hazard (Clark, 1987, Bruijnzeel, 2004, Bradshaw et al., 2007, Van Dijk et al., 2009), although it should be cautioned that the results could also be explained by other confounding factors (e.g. agricultural development, urbanisation) that may co-occur alongside changes in forest cover. It should also be noted that although the Global Forest Change dataset is the most advanced product currently available of global tree canopy, the period of this dataset does not fully overlap the streamflow records, which is the main limitation of this analysis.

Importantly, for almost all datasets considered, and regardless of whether the stations were filtered by the presence of dams, catchment area, or forest cover changes, there were more stations exhibiting significant decreasing

trends than increasing trends. The results of our global analysis were therefore generally consistent with Kundzewicz et al. (2004), who also showed that the number of stations showing significant increasing trends was found to be smaller than the number of stations showing significant decreasing trends. The exception of our analysis was for the "no dam" dataset for the smallest two quintiles by catchment size, where there were more stations exhibiting increasing than decreasing trends. This result was inconsistent with what was found in Milly et al. (2002), although that study focused only in the frequency of extreme events (i.e. discharges exceeding 100-year levels), whereas our study focused on annual maximum streamflow and thus includes relatively moderate 'flood' events as well as rarer events. There are multiple factors that might explain the effect of catchment size on the trend results; for instance, larger catchments typically are affected by longer time-scale rainfall events, and thus the climatic forcings may be different compared to smaller catchments (Westra et al., 2014). Furthermore, catchment size may be associated with other anthropogenic modifications in the catchment, such as the presence of agriculture, urbanisation, and the construction of a range of hydraulic infrastructure not included in the large dams database.

Despite potential concerns about data quality, one interesting pattern to emerge was that detected changes in annual maximum streamflow are inconsistent with the evidence of trends in precipitation. At global scale, annual maximum precipitation intensities were found to have increased (Min et al., 2011) and a large-scale increasing pattern in extreme precipitation was detected (Lehmann et al., 2015), with North America experiencing more increasing trends than decreasing trends in annual maximum precipitation (Westra et al., 2013). These precipitation-based results therefore appear to be inconsistent with the trend results for floods found in our analysis, and indicate the potentially important role of changes in catchment conditions and river morphology to changes in streamflow regimes (Hall et al., 2014, Merz et al., 2012a). Further research is needed to quantify the contribution of catchment condition to the rainfall-runoff relationship at global and regional scales, including investigation of changes in other dimensions of flooding, such as their duration, volume, and intensity.

Finally, the changes in the flood hazard as assessed in this study do not explain observed increases in flood losses (Kundzewicz et al., 2013, Mills, 2005) or in the number of reported events (Munich Re, 2015, Swiss Re, 2015). In our results, there were more decreasing trends than increasing trends in almost all cases, regardless of the dataset used, whether catchments were filtered by climatic condition, presence of dams, or forest cover loss. This indicates that other factors contributing to flood risks (i.e. exposure and vulnerability), are likely to contribute a higher share toward the rise of flood losses at a global scale.

Acknowledgement

We thank the Global Runoff Data Centre (GRDC) for providing the observed streamflow data. Mr Hong Xuan Do receives financial support from the Australia Award Scholarship (AAS). Dr Westra's time was supported by Australian Research Council Discovery project DP150100411. The authors also wish to thank two anonymous reviewers for their constructive comments and suggestions.

Chapter 4. Global-Scale Prediction of Flood Timing Using Atmospheric Reanalysis (Paper 3)

Statement of Authorship

Title of Paper	Global-scale prediction of flood timing using atmospheric reanalysis
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Do, H.X., Westra, S., Leonard, M., Gudmundsson, L., 2019. Global-scale prediction of flood timing using atmospheric reanalysis. Water Resources Research, (submitted).

Principal Author

Name of Principal Author (Candidate)	Hong Xuan Do
Contribution to the Paper	Designed the experiments, data reprocessed, conducted statistical analyses, analysed results, wrote manuscript and acted as corresponding author.
Overall percentage (%)	85
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 07/02/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Seth Westra
Contribution to the Paper	Suggest scope of study, helped to design experiments and interpret results, provided feedbacks on manuscripts and responses to reviewers.
Signature	Date 15/02/19

Name of Co-Author	Michael Leonard
Contribution to the Paper	Suggest scope of study, helped to design experiments and interpret results, provided feedbacks on manuscripts and responses to reviewers.
Signature	Date 19/02/2019

Name of Co-Author	Lukas Gudmundsson
Contribution to the Paper	Suggest scope of study, helped to design experiments, provided feedbacks on manuscripts and responses to reviewers.
Signature	Date 11/07/2019

Abstract

Flood seasonality is a useful indicator to study the interaction between atmospheric and catchment processes in generating floods. This paper presents a global assessment of the seasonal timing of floods for 7,894 gauging locations across the globe over a common period from 1981 to 2010. The averaged ordinal date of annual maximum streamflow is then estimated for ungauged locations following a two-stage prediction scheme. The first stage of the prediction scheme identifies regions that share a common flood producing mechanism by analysing the observation of flood timing with respect to seven climate predictors representing precipitation timing and snow melt dynamics, which are derived from a global climate reanalysis dataset. The homogeneous regions in terms of the dominant flood generation process are generalised in the second stage of the prediction model through a rule-based classification. The classification partitions the world into ten hydro-climate classes, where each class has flood timing predicted using the most relevant climate predictor. Using this relatively simple and interpretable model structure, a global mean absolute error of 31.9 days was achieved, whilst maintaining consistency across large regions in the estimates of timing. Potential applications of the developed map include benchmarking the performance of global hydrological models and assessing the impact of climate change on the timing of floods.

4.1 Introduction

The seasonal timing of flood events is a useful indicator of how atmospheric processes interact with the local catchment, with recent papers showing the relevance of intense precipitation, snow melt and rain-on-snow events as mechanisms driving the flood timing (Parajka et al., 2010, Blöschl et al., 2017, Villarini, 2016, Hall and Blöschl, 2018, Iliopoulou et al., 2019). An understanding of flood timing provides useful insights at many scales: (i) globally – because of the considerable attention devoted to the development of global hydrological models (Bierkens, 2015, Wood et al., 2011, Bierkens et al., 2015), and the need to reconcile patterns of non-stationarity in climatic drivers such as rainfall (Westra et al., 2013, Westra et al., 2014, Sharma et al., 2018) with those observed in streamflow (Do et al., 2017, Hodgkins et al., 2017, Gudmundsson et al., 2019, Gudmundsson et al., 2017); (ii) regionally – for analyses of flood frequency within homogeneous regions and for detection/attribution of historical changes in flooding (Villarini, 2016, Cunderlik et al., 2004); and (iii) locally – to assist understanding of flood mechanisms, as required by decision makers in designing strategies for flood prevention, mitigation, and response (Dhakal et al., 2015, Ward et al., 2015).

There have been many studies of flood magnitude and frequency characteristics: (i) at global (Asadieh et al., 2016, Dankers et al., 2014, Wasko and Sharma, 2017, Woldemeskel and Sharma, 2016, Do et al., 2017, Hodgkins et al., 2017); (ii) continental (Ishak et al., 2013, Hall et al., 2014, Gudmundsson et al., 2012a, Mediero et al., 2015, Alfieri et al., 2015, Parajka et al., 2010, Mallakpour and Villarini, 2015); and (iii) national scales (Beurton and Thieken, 2009, Stevens et al., 2016, Slater and Villarini, 2016, Merz et al., 2018, Burn and Whitfield, 2016), but comparatively fewer and more recent studies of flood timing (Cunderlik and Ouarda, 2009, Burn and Whitfield, 2016, Blöschl et al., 2017, Ye et al., 2017, Berghuijs et al., 2016, Villarini, 2016, Dettinger and Diaz, 2000, Hall and Blöschl, 2018). Interestingly, most of the studies of flood timing find unique information on the atmospheric mechanisms that cause floods, where unlike indicators of flood magnitude or frequency, the average timing of floods is relatively independent of characteristics such as catchment size, shape, and elevation

(Hall and Blöschl, 2018), or human influences including land use change and river regulation (Villarini, 2016).

Of the studies that have considered flood timing, most have focused on Europe and North America, so that a global perspective of when and why floods occur at different times of the year is not yet available. To develop this global perspective, it is essential to expand the assessment of flood timing to other continents (e.g. Australia, South America, Asia, Africa) using consistent datasets and analysis methodology. One possibility is to simulate runoff and extract information of flood timing through the use of global hydrological models (Lee et al., 2015) forced with global reanalysis climate. To our knowledge, Lee et al. (2015) is the only model-based study to produce a global map of the peak flow season (defined as the consecutive three month period with the highest number of events above a threshold of streamflow volume), whereas model-based studies of timing of annual maximum flow are not yet available. An alternative possibility is to estimate flood timing using available observational datasets from across the globe and construct a data-driven model to infer flood timing at locations without streamflow observations. In addition to providing meaningful information in its own right, such an approach would provide a useful point of comparison for any subsequent model-derived maps of flood timing.

The spatial variation of the dominant mechanisms in flood generation, however, poses a challenge to predicting flood timing for ungauged locations. Heavy rainfall is one of the most common sources of flooding, as the catchment rapidly saturates due to receiving a significant amount of precipitation (Kozlowski, 1984). However, many studies have shown that other factors also play an important role in the flood generating processes, including antecedent soil moisture (Ivancic and Shaw, 2015, Wasko and Sharma, 2017, Ye et al., 2017, Bennett et al., 2018) and snowmelt dynamics (Parajka et al., 2010, Blöschl et al., 2017, Mediero et al., 2015, Berghuijs et al., 2016). Flooding in arid regions or very large catchments may be more sensitive to the total amount of rainfall over long periods (up to months) rather than short duration rainfall events (Johnson et al., 2016, Ingle Smith, 1999, Marengo, 2006, Pathiraja et al., 2012), and thus the long-term total

precipitation also needs to be taken into account. A reliable model for flood timing, therefore, must possess the capacity to define homogeneous regions in terms of flood generation processes, which will then be used as the prediction basis.

The recent publication of a global archive of over 30,000 streamflow gauges (GSIM; Do et al., 2018b, Gudmundsson et al., 2018b) provides a unique opportunity to explore many aspects of streamflow characteristics at the global scale, including flood timing. The main aim of this study is to use this resource, combined with an atmospheric reanalysis dataset, to develop a data-driven model to infer flood timing at both gauged and ungauged regions across the globe. Specifically, global seasonality of flood timing is first evaluated across all GSIM stations with sufficient data. Observations of flood timing are then analysed with respect to seven climate predictors derived from the ERA-Interim dataset to identify potential flood producing mechanisms, and the single predictor best suited to explain and predict flood occurrence is identified at each location. The regional consistency between flood timing and the most relevant predictor can then be generalised to all gauged and ungauged locations using a rule-based classification system according to ten hydro-climate classes. The remainder of this paper is structured as follows. Section 4.2 provides an overview of the data and methods that were used to assess the seasonal timing of floods and the prediction scheme development. The results are reported in Section 4.3 together with discussions about the performance of the prediction scheme. Finally, Section 4.4 summarises the key findings and highlights potential application of the proposed prediction scheme.

4.2 Data and methods

This section summarizes the workflow for global prediction of flood timing (Figure 4-1), including input variables, observational analyses, and the prediction scheme using a rule-based classification system. The datasets used in this study are presented in section 4.2.1, followed by descriptions about observations and predictors of flood timing (section 4.2.2) and the

development of a prediction scheme to derive a global map of flood timing (section 4.2.3).

4.2.1 Datasets

The Global Streamflow Indices and Metadata (GSIM) archive contains streamflow indices from more than 30,000 stations across the globe (Do et al., 2018b, Gudmundsson et al., 2018b). To establish a compromise between data quality and availability, only stations classified with a 'useful' homogeneity class (Gudmundsson et al., 2018b) are used to ensure that stations with potentially spurious step changes are excluded. A threshold of at least 20 yearly data points available during the 1981-2010 common period (with each year having at least 350 days of reliable records) was used to select streamflow gauges with sufficient data to minimise the influence of inter-annual and inter-decadal variability, while maintaining a relatively large sample for a global scale investigation. This filtering process identified 9,560 viable stations, of which a further 76 stations were removed due to unavailability of catchment area information. To mitigate the influence of large-scale climate gradients as well as routing effects and catchment processes, an approach of previous global-scale reconstruction studies was adopted (van Dijk et al., 2013, Beck et al., 2015) to limit stations to those with catchment area less than 10,000km^2 (1,226 stations were removed), approximately the size of a one-degree longitude/latitude grid cell. Finally, 364 stations that fall outside of ERA-Interim land regions were also removed, as this data is required to develop the prediction scheme. The outcome of this filtering process was the identification of a final subset of 7,894 stations to be used for this study.

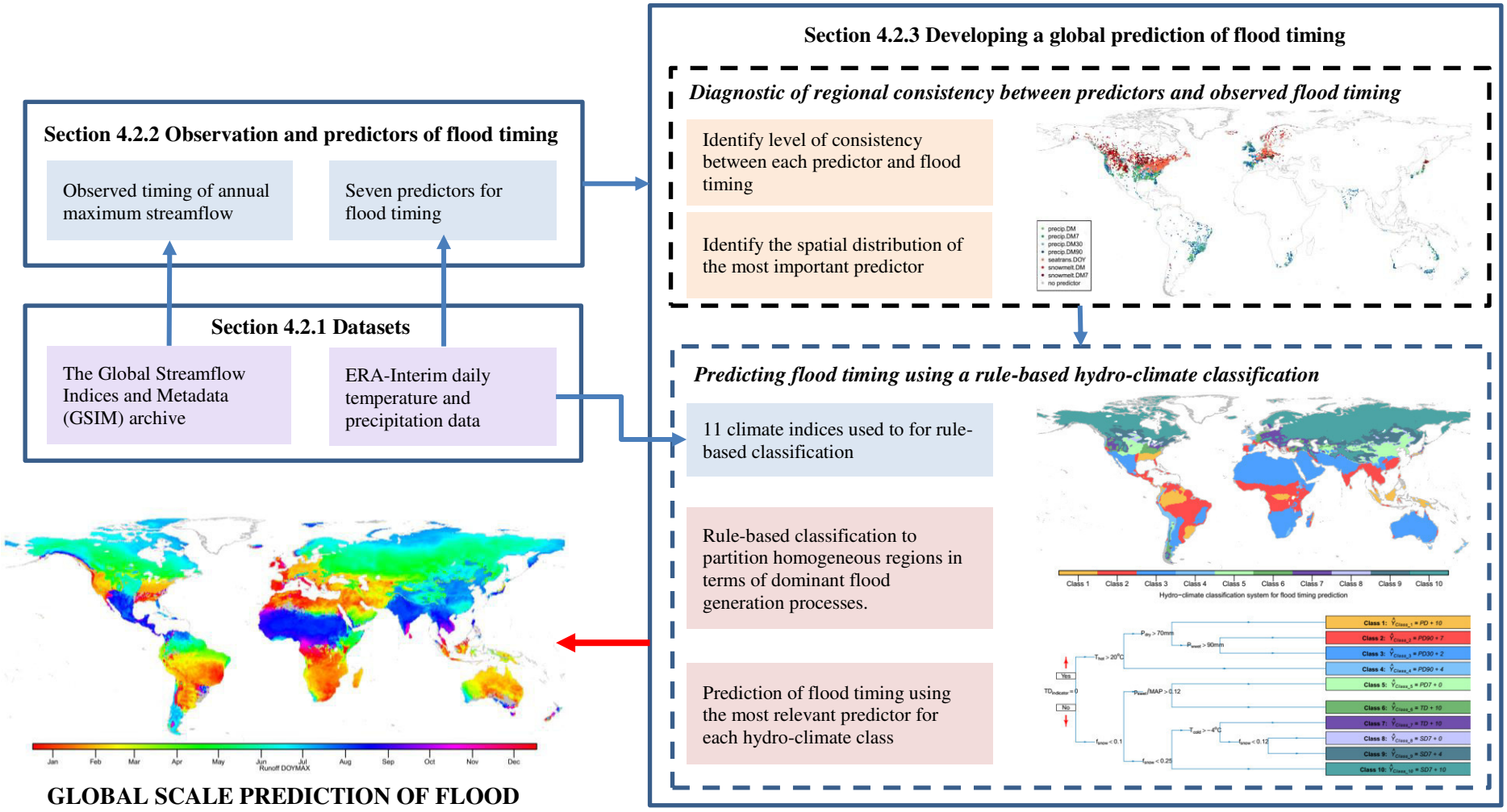


Figure 4-1. Flow chart to make global prediction of flood timing using GSIM and ERA-Interim datasets.

To represent global observation of atmospheric forcing, the ERA-Interim dataset was used over the same 1981-2010 period (Dee et al., 2011). Regridded daily temperature and precipitation data products at 0.5-degree resolution were retrieved directly from the European Centre for Medium-Range Weather Forecasts data portal. The land-sea mask from ERA-Interim was used to keep only values over land regions (except Greenland and Antarctica, which were excluded). Time series at monthly and annual resolutions were aggregated from original daily time series. Reported streamflow gauge coordinates (Do et al., 2018a) were used to identify corresponding grid-cells from the global climate dataset and extract information of both precipitation and temperature for each streamflow station. This dataset covered all the global land regions, providing the capacity to make predictions of flood timing across the globe.

4.2.2 Observations and predictors of flood timing

4.2.2.1 Observations of flood timing

The ordinal day of annual maximum streamflow (DOYMAX index, available in GSIM (Gudmundsson et al., 2018a)) was selected as the indicator of flood seasonality. Circular statistics (Mardia and Jupp, 2009) were used to assess the seasonality of historical flood timing , with further details provided in the supplementary material. The circular-mean value of each DOYMAX time series over the period from 1981-2010 was used as the observed timing of flood seasonality for each stream gauge. A concentration statistic (R) of each DOYMAX time series was also calculated to represent the strength of the seasonality, where $R = 0$ indicates that flood occurrence dates were spread evenly throughout the year, and $R = 1$ indicates that all flooding events occur on the same ordinal day across all years. Note that a low value of the flood timing concentration index R does not always correspond to low levels of seasonality and could reflect other complex flood timing distributions (e.g. reflective symmetric bi-modal, or asymmetric unimodal), which is beyond the scope of our investigation. Stations with non-seasonal flood timing were identified through a circular Kuiper's test, which evaluates whether the time series is circularly uniform. Only stations for which the null hypothesis of circular uniformity is rejected at the 10% confidence level (i.e. those stations

that have statistically significant seasonality) were considered as input for the prediction of flood timing.

Figure 4-2 shows examples of calculating the mean and concentration of flood timing. The left panel illustrates a location where flood events can occur at any time of the year. The hypothesis of uniformity was not rejected at the 10% significance level in this case, suggesting the absence of evidence of flood seasonality. The right panel provides an example of seasonality where all flood events occur between November to April and the majority of the events fall in January and February.

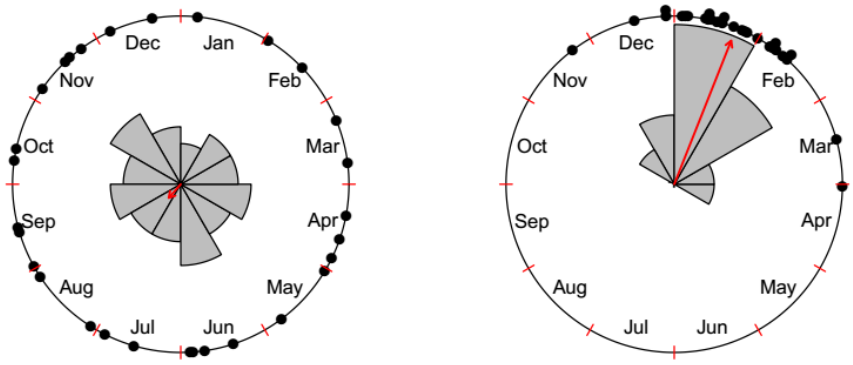


Figure 4-2. Example of a station that does not have evidence to reject the null-hypothesis of uniformity in a circular time series (left panel; the east branch of Cann River located in Victoria, Australia), and a station that has evidence to reject the uniformity hypothesis (right panel; Los Sosa River located in Entre Rios Province, Argentina). Grey areas represent the density of maximum streamflow events distributed across 12 months of the year. The direction of the red arrow represents the average timing, whereas the length of the arrow illustrates the concentration index of the record (R value).

4.2.2.2 Predictors of flood timing

This section presents the seven climate predictors of flood timing, which were considered in this analysis. To ensure global availability for the prediction, daily precipitation and temperature data at each grid point of the ERA-Interim datasets were used to derive the identified predictors. Each predictor is the circular mean value of occurrence date of one hypothesised flood producing process over the 1981-2010 period. The seasonality assessment using the circular uniformity hypothesis was also applied to these seven predictors. Only grid cells where the null hypothesis of circular uniformity was rejected at the 10% level are considered, while a missing value was assigned for other grid cells. The seven climate predictors are

divided into three groups based on the key flood generating processes that they represent.

The first group of predictors focused on short-term precipitation and reflect the hypothesis that extreme precipitation events are the primary mechanism driving large streamflow events. Based on the contributing areas of the gauging stations, it is estimated that all stations in the final subset (catchment area less than 10,000km^2) had times of concentration of seven days or less – based on the Pilgrim McDermott formula (Pilgrim et al., 1987), suggesting heavy precipitation events spanning a period of seven days or less is appropriate to represent this mechanism. There are two variables considered in this group: (i) date of peak daily precipitation in each calendar year (PD), which represents flood produced by the single largest precipitation event, and (ii) date of peak 7-day precipitation in each calendar year (PD7), which represents flood produced by the largest series of precipitation events. To extract the PD7 variable, a backward-moving window of seven days was applied to each day of the year, and the day of maximum value (comprising the total rainfall depth on that day and the six prior days) was recorded for each calendar year.

The second group of predictors focus on long-term precipitation and reflect the hypothesis that long-term catchment wetness and antecedent moisture conditions play a key role in the flood generating process. There are two variables considered in this group: (i) the date of peak 30-day precipitation in each calendar year (PD30), which represents the hypothesis that the peak discharge occurs when the drainage area is relatively wet, and (ii) the date of peak 90-day precipitation in each calendar year (PD90), which represents the hypothesis that the timing of peak discharge occurs toward the end of a wet season, where significant build-up of catchment moisture will have occurred. The calculation process for PD30 and PD90 was similar to PD7 variable, but with the backward-moving window set at 30 days and 90 days respectively.

The third and final group of predictors focus on snowmelt processes and are designed to provide an indicator of snowmelt or rain-on-snow processes. The first predictor in this group is the date of seasonal transition from

snowfall to rainfall in precipitation (TD), which is defined as the first day that the surface air temperature rises above 0°C after having been below 0°C for at least seven consecutive days (Blöschl et al., 2017). To represent a more sophisticated indicator for snowmelt events, a simple degree-day method (Woods, 2009, Hock, 2003) was used to simulate snow-dynamics (see supplementary material for detail methodology). This led to two predictors derived from simulated snowmelt contribution: (i) date of peak value of daily snowmelt or rain-on-snow (SD) and (ii) date of the peak value of 7-day snowmelt or rain-on-snow (SD7; here a backward-moving window of 7 days was used to calculate the time series of total snowmelt or rain-on-snow amount). To mask out locations where there was an absence of significant contribution of snowmelt to flood generation, additional constraints were applied to snowmelt predictors. For the TD predictor, locations where this variable cannot be identified for more than 70% of the years were assigned a missing value. Missing values were also assigned to SD and SD7 predictors across locations where less than 10% of precipitation falls as snow.

The availability of chosen climate predictors across the globe is shown in Figure 4-3. The constraining criteria for snowmelt predictors imply that these predictors are mostly available in high latitude regions in the northern hemisphere, and in some mountainous areas in the southern hemisphere such as the Andes in South America and the Southern Alps in New Zealand. Due to non-seasonality of the selected predictors (i.e. where the circular Kuiper test does not reject the uniformity hypothesis), some areas, mostly desert regions, do not have any available predictors, such as in the interior of southern Australia, the south-eastern part of the Arabian Peninsula, or the Uruguay River. Furthermore, only snowmelt predictors are available for many grid cells across the Appalachian Mountains in North America, Eastern Europe, central Kazakhstan, and northern Africa as a result of the lack of rainfall seasonality in these regions. Detailed maps of the timing and seasonality of each predictor are provided in the supplementary materials (Figures P4-1 to P4-7).

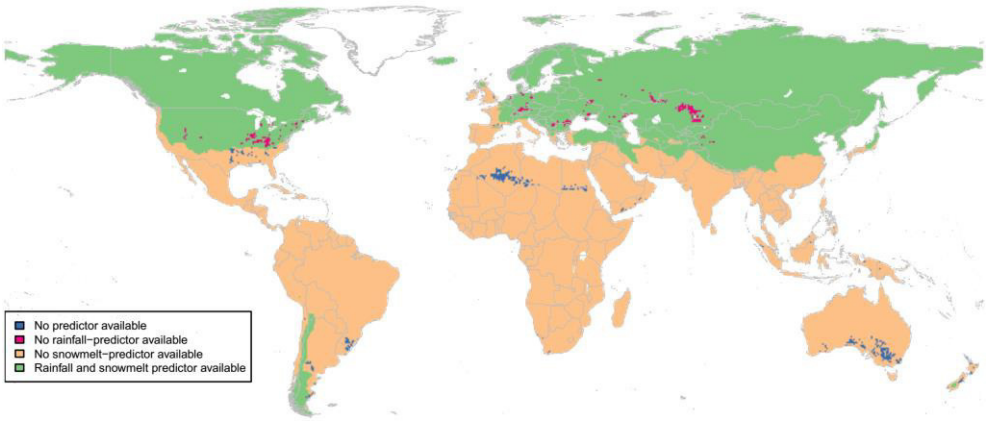


Figure 4-3. Map of data availability for the seven predictors. Predictors were divided into two categories: (1) Rainfall-predictors comprising PD, PD7, PD30 and PD90 and (2) Snowmelt-predictors comprising TD, SD and SD7. Unavailability may be due to no data being available (for snowmelt-base predictors only) or where the circular uniformity hypothesis was not rejected at the 10% level (for all predictors).

4.2.3 Developing a global prediction of flood timing

At the global scale, it has been shown previously that the mechanism dominating flood occurrence varies significantly in many regions (Berghuijs et al., 2016, Blöschl et al., 2017), and thus a reliable prediction for flood timing must adequately reflect this spatial variation. To facilitate this requirement, the present study proposes a two-stage prediction model, in which the first stage (sub-section 4.2.3.1) aims to define homogeneous regions in terms of the dominant flood producing mechanism. In the second stage (sub-section 4.2.3.2), the defined homogeneous regions are generalised across the globe through a classification scheme, in which prediction of flood timing is made for each class by a linear function of the most relevant predictor. The global prediction for flood timing was then obtained by applying the classification system and the linear functions to all land locations, including ungauged regions.

4.2.3.1 Diagnostic of regional consistency between predictors and observed flood timing

This section describes the first stage of the flood timing prediction scheme, aiming to define regional patterns of dominant flood generation mechanisms from observational data. The discrepancy between the average ordinal dates of predictors and annual maximum streamflow events was first calculated to identify the climate variable with the closest match at each

location across 7,040 stations. The level of consistency between flood timing and the predictor with the closest match was assessed by grouping stations into five categories based on the magnitude of discrepancy, as outlined in Table 4-1. The spatial distribution of the single predictor with the highest level of consistency to flood timing at each gauged location was then used to represent homogeneous regions in terms of the dominant flood generation process.

Table 4-1. Description of the five consistency categories between flood timing and a single predictor

Category	Description
High consistency	Discrepancy between (±)15 days (i.e. within a month)
Medium consistency	Discrepancy between (±)16 and (±)45 days (i.e. within a season)
Low consistency	Discrepancy between (±)46 and (±)75 days
Inconsistency	Discrepancy is outside of [-75, +75] range
No data available	Predictor data is not available at the reported coordinates of the streamflow station due to seasonal uniformity of the time series

4.2.3.2 Predicting flood timing using a rule-based hydro-climate classification

In the second stage of the prediction scheme, the observed homogeneous regions were generalised across the globe through a rule-based classification system, which used 11 climate indices (derived from ERA dataset and summarised in Table 4-2) as separating variables. The classification scheme in this study has a similar structure to that of a classification tree, which is a binary tree with nodes defined by simple splitting rules applied to a set of input variables and corresponding thresholds (e.g. at the root node, all stations are divided into two groups by a decision rule 'fraction of precipitation falls as snow is less than 0.1'). However, each leaf of the tree (i.e. terminal node or hydro-climate class in the context of this study) provides a prediction of flood timing through a linear function of one of the seven climate predictors rather than being simply assigned a class.

To develop this classification scheme, one possible option is to apply machine learning techniques such as recursive binary splitting together with a greedy pruning algorithm (see Cannon (2012) for example) on available datasets. However, we decided to construct the model in a semi-automated manner to ensure that the final hydro-climate classification system is

physically interpretable while regional patterns of predictors that best explain the occurrence of flood are retained.

Table 4-2. Climate indices that were used as input for the rule-based hydro-climate classification.

Index	Description
MAP	mean annual precipitation (m)
MAT	mean annual temperature (°C)
T_{hot}	temperature of the hottest month (°C)
T_{cold}	temperature of the coldest month (°C)
P_{dry}	precipitation of the driest month (m)
P_{sdry}	precipitation of the driest month in summer(*) (m)
P_{wdry}	precipitation of the driest month in winter(*) (m)
P_{swet}	precipitation of the wettest month in summer(*) (m)
P_{wwet}	precipitation of the wettest month in winter(*) (m)
f_{snow}	fraction of precipitation falling as snow (from 0 to 1). Daily precipitation is assumed to fall as rainfall when $T > 0$
$TD_{indicator}$	Binary variable (0/1) indicates whether transition time from snowfall to rainfall can be reliably identified (i.e. at least 70% of the years have a temperature rise from below to exceed 0°C).

(*): summer (winter) is defined as the warmer (cooler) six-month period of October – March and April – September for each respective hemisphere.

At each non-terminal node n, a specific climate index (selected from 11 climate variables and denoted here by C_n) and corresponding thresholds were manually selected to divide the world into sub-regions. The selection of a climate index C_n was based on visually matching the spatial variations of all climate indices (see supplementary Figures C4-1 to C4-11) to the regional consistency between predictors and observed flood timing (results of the method presented in Section 4.2.3.1). This procedure was repeated until appropriate boundaries were drawn to divide the world into hydro-climate classes.

We define the terminal node of the partitioning scheme (hydro-climate class R_j, where j is the index over all classes) to be a homogenous 'region' that shares a common group of flood timing predictors. At each terminal node R_j, the timing of the flood (denoted by Y_{R_j}) is then predicted by adding a lag-day (denoted by γ_{R_j}) to the value of the best climate predictor (denoted by X_{R_j}, which is one of the seven climate predictors that was defined in section 4.2.2.2). For a specific hydro-climate class R_j, the prediction of flood timing ($\hat{Y}_{R_j}$) was made using a linear equation:

$$\hat{Y}_{R_j} = X_{R_j} + \gamma_{R_j} \qquad j = \{1,2,\dots,J\}$$

The central idea of this prediction scheme is that regions with the same hydro-climate class (R_j) are likely to have floods, on average, occurring γ_{R_j} days after the occurrence of the hypothesised mechanism (γ_{R_j} is bounded between 0 and +20 days). For example, if the peak daily precipitation (PD) was identified as the most suitable predictor for hydro-climate class R_j, the flood timing at each station in this class is predicted by adding a constant γ_{R_j} to the date of the peak daily precipitation.

The best predictor and corresponding lag-day for each hydro-climate class was determined following an automated optimisation. The objective of the optimisation was to (i) minimise error between predicted and observed flood timing and (ii) maximise the proportion of locations that have available data for the predictor. The former criterion represents the predictive ability while the latter indicates the ability to correctly identify regions with common flood generating processes of the prediction scheme. The objective function used an adjusted mean absolute error:

$$AMAE_{R_j} = \frac{1}{P_{R_j}^2} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{N} \qquad i \in R_j, \quad j = 1,2,\dots J$$

where $AMAE_{R_j}$ is the adjusted mean absolute error for region R_j, N is the number of stations located in hydro-climate class R_j, $\hat{y}_i$ is the prediction while y_i is the equivalent observation of flood timing for a specific site i within hydro-climate class R_j, and P_{R_j} is the proportion of locations in hydro-climate class R_j having available data for predictors (ranging from 0 to 1). This metric was used to penalise predictors that are unavailable for many locations within a specific hydro-climate class (the square value emphasises the importance of this metric). The predictor X_{R_j} and value of γ_{R_j} that minimised the error for a given climate class were selected for each terminal node of the prediction scheme.

The tree-based flood timing prediction model, calibrated to the seasonal flood timing observed across the 7,040 stations, was then applied to all land grid cells of ERA-Interim dataset to derive a map of flood timing.

4.3 Results and discussion

4.3.1 Seasonality characteristics of flood at the global scale

Figure 4-4 provides an overview of flood seasonality at gauged locations at the global scale. The top panel illustrates the average timing of floods for the 1981-2010 period, while the lower panel shows the flood timing concentration R. Stations that exhibit uniformity in the records are highlighted as red dots in the lower panel. There is a clear regional association in the timing of flood occurrence, of which the patterns over North America and Europe concur with prior studies (Villarini, 2016, Blöschl et al., 2017, Hall and Blöschl, 2018, Burn and Whitfield, 2016).

With improved data coverage, this analysis also provides a regional perspective of flood timing over Asia (the majority of stations is located in Japan and India) and the southern hemisphere (the majority of stations is located in Brazil). In Asia, high latitude regions have floods characterised by spring snowmelt while the rest of this continent is dominated by summer to autumn floods, typically related to the monsoon season and tropical storms. In the southern hemisphere, there is a clear transition of flood timing in the latitudinal direction, strongly suggesting the influence of large-scale atmospheric features on the regional hydrological cycle. Due to the limited availability of snowmelt processes in the southern hemisphere (only significant in some mountainous areas as discussed in section 4.2.2.2), the rainfall regime and its interaction with catchment soil moisture conditions are more likely to be the key flood generation mechanisms across these regions.

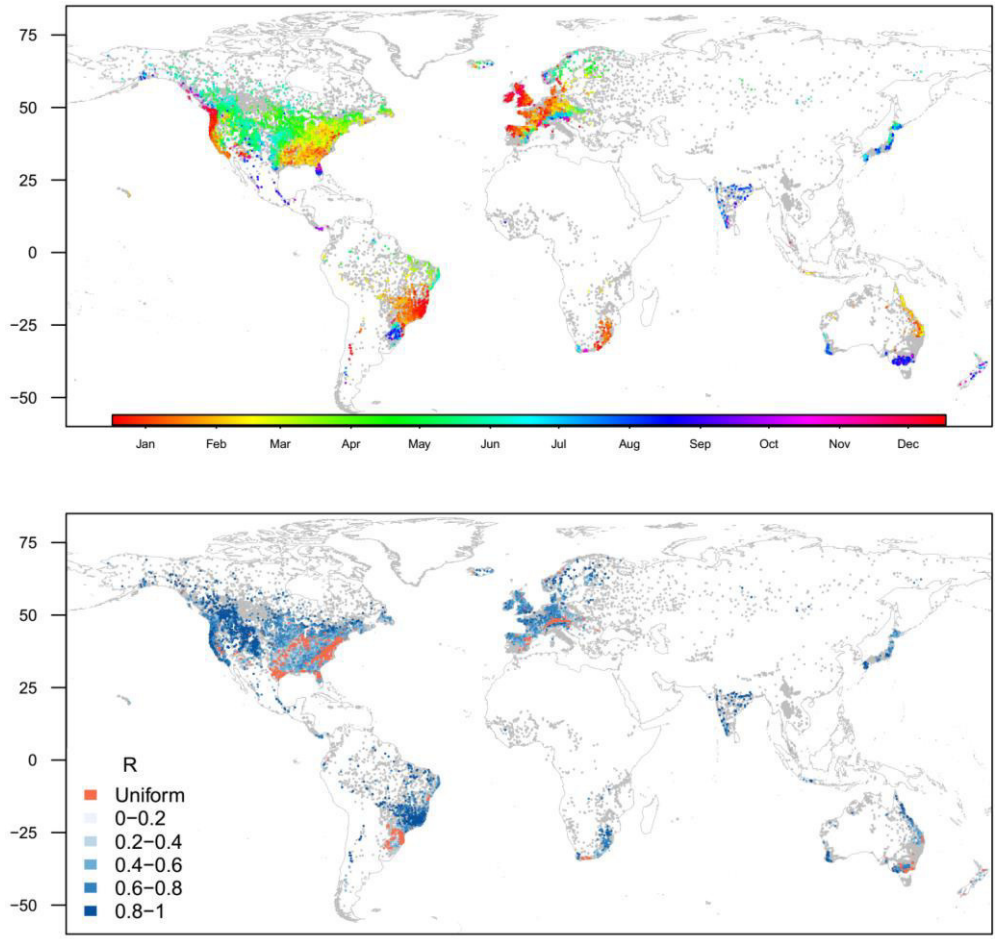


Figure 4-4. Seasonality of flood occurrence across 7,894 GSIM stations fulfilling the quality control criteria for the period 1981-2010. Top panel: average flood timing; colour points represent long-term-mean value. Lower panel: concentration index (R) of flood timing (values range from 0 to 1); red dots represent records with uniformity hypothesis was not rejected at the 10% significance (854 stations). In both panels: grey dots represent GSIM stations that were removed prior to this analysis.

The strength of the seasonal cycle (lower panel of Figure 4-4) demonstrates a high level of heterogeneity in the seasonal signal. There are several clusters of stations showing uniformity due to the influence of climate-related processes that have been documented in previous studies. For instance, the east of U.S. is subject to a range of flood generation processes occurring throughout the year such as tropical and extratropical storms (Smith et al., 2011), or snowmelt dynamics (Villarini and Smith, 2010). European stations located at the foothill of mountainous areas tend to be influenced by a mix of spring-snowmelt, rainfall events and/or glacier melting in summer (Hall and Blöschl, 2018). The southern coast of south-eastern Australia has frequent rainfall in winter, but heavier summer precipitation is also possible due to convective activity. The combined effect of extreme rainfall and

antecedent soil moisture is a likely reason for uniformity in flood timing records across this region (Leonard et al., 2008), particularly where soil moisture conditions are counter-cyclical with heavy rainfall (e.g. the most intense rainfall may occur during summer due to convective processes, but on average the soils tend to be the wettest during the winter). Lastly, the south of Brazil is characterised by a non-defined rainy season due to the combined influence of cold fronts, thunderstorms, and tropical cyclones which make rainfall-induced floods occurring throughout the year (Rao and Hada, 1990, Teixeira and Satyamurty, 2011). Ultimately of the 7,894 selected records, the uniformity hypothesis was rejected for 7,040 locations, and this subset of stations that exhibited significant seasonality in flood timing represents the final subset used for the prediction of flood timing.

4.3.2 Distribution of predictors with the least discrepancy to flood timing

The distribution of the 'best climate predictor' for the globe is provided in Figure 4-5 (regional maps for areas with a high density of stations are provided in supplementary Figure S4-1). An interesting pattern observed through this analysis is the high level of spatial clustering in the distribution of predictors having the least discrepancy to flood timing, suggesting the existence of homogeneous regions in terms of flood generating processes.

In regions above 35°N where snowmelt plays an important role in flood generation, there are clear regional patterns regarding the most important predictor of flood timing. In particular, the rainfall-dominant predictors (i.e. PD, PD7, PD30, and PD90) are generally the most suitable to explain flood occurrences on the western coastline of North America and Western Europe (including the UK), while snowmelt-dominant predictors (i.e. TD, SD and SD7 which usually occur in spring) are generally most suitable in the north-central and the north-east of the U.S., most of Canada, Central and North-Eastern Europe, North Eurasia, and Scandinavia. These findings are generally consistent with previous studies (Ye et al., 2017, Berghuijs et al., 2016, Villarini and Smith, 2010, Cunderlik and Ouarda, 2009, Burn and Whitfield, 2016, Mediero et al., 2015, Hall et al., 2014).

Focusing on regions with no snowmelt-based predictors (i.e. below 35°N), short-term precipitation predictors (PD and PD7) generally have the closest match with the timing of floods in the south-eastern US, northern Australia, and both the eastern and southern regions of Brazil, where previous studies have shown the importance of thunderstorm activities or tropical cyclones in flood generation (Villarini, 2016, Bradley and Smith, 1994, Villarini et al., 2014, Stevenson and Schumacher, 2014, Ávila et al., 2016). On the other hand, long-term precipitation predictors (PD30 and PD90) have the highest consistency with flood timing in central Brazil and southern Australia, while other regions show a mixture between these two groups.

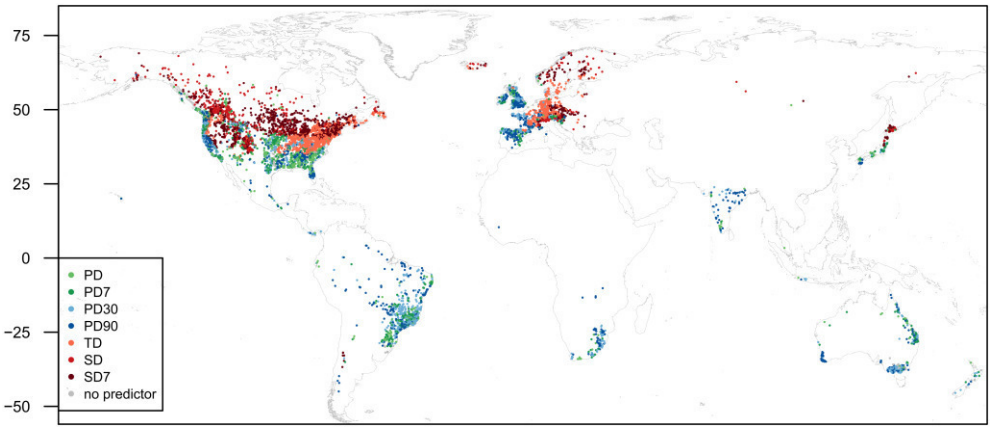


Figure 4-5. Global map of single predictor with smallest discrepancy to flood timing across 7,040 stations that exhibit seasonality in flood timing. The green colours indicate the short-precipitation predictor (PD and PD7), blue colours represent the long-precipitation predictors (PD30 and PD90) and the red colours represent the snowmelt-base predictors (TD, SD and SD7). There are 63 stations with no data available for predictors. These stations are plotted in the grey colour.

This comparison shows two of the main challenges for predicting flood timing at the global scale. Firstly, within relatively small geographic areas, such as the US Rocky Mountains or the Alpine region in Europe, there is large variability in the identified predictor, which reflects the complexity of flood formation factors (snowmelt, soil moisture state of the catchment, and different types of precipitation) across these regions (Parajka et al., 2010, Berghuijs et al., 2016). Secondly, many locations also show a high correlation between predictors (e.g. the average timing of short-term precipitation and long-term precipitation being in the same month, see supplementary Figure S4-2), and this feature creates noise in determining the most important predictor. Nevertheless, the spatial patterns shown in Figure 4-5 indicate the

utility of the climate predictors to identify different flood-timing mechanisms at the regional scale.

The level of consistency between flood timing and available predictors (i.e. the discrepancy, in number of days, between flood timing and available predictors as defined in Table 4-1) was also analysed to evaluate the appropriateness of using these predictors for estimating flood timing. At the continental scale (Figure 4-6), all precipitation-based predictors generally have a good level of consistency in Asia, Africa, and South America, with more than 70% of stations exhibiting high or medium consistency with flood timing. In Oceania (of which the majority of stations are in Australia), flood timing is most consistent with long-term precipitation predictors, as both PD30 and PD90 have more than 60% of stations exhibiting high or medium consistency. In North America and Europe, where snowmelt-related processes are a key flood-producing mechanism, the percentage of stations showing high or medium consistency between precipitation-based predictors and flood timing is lower than the other continental regions; however, this is supplemented by snowmelt predictors, which have high and medium consistency for approximately 25-40% of stations.

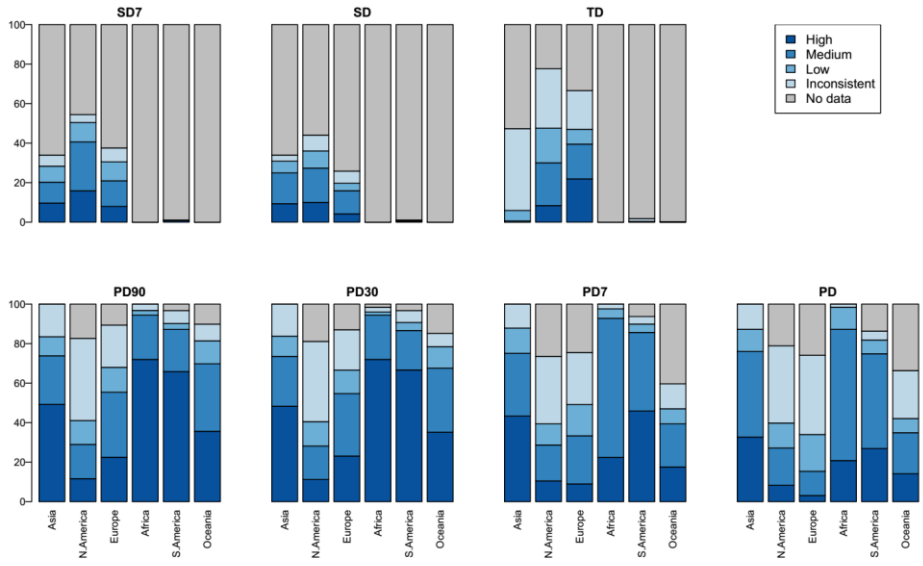


Figure 4-6. Consistency between flood timing and individual predictors (top panels: snowmelt-based predictor; bottom panels: rainfall-based predictors), based on definitions in Table 4-1. Each bar chart illustrates the percentage of stations allocated into five consistency categories for one predictor across the six considered regions.

The level of consistency between flood timing and the single most important predictor across the 7,040 stations was also assessed (showed in

Table 4-3), suggesting generally high consistency at the global scale with the percentage of stations having high and medium levels of consistency being 50.9% and 31.8% respectively. This pattern is also evident at the continental level, with the percentage of locations showing high or medium consistency levels ranging from 72% (Oceania) to 97% (Africa). These results indicate the potential of using the proposed indices to predict flood timing, which could result in a model with up to 80% of locations having a prediction error of less than 46 days (i.e. the predicted and observed flood timing will fall within the same season).

Table 4-3. Number of stations grouped by five consistency categories at regional and global scales.

Continents	Level of consistency					Total
	High	Medium	Low	Inconsistent	No data	
Asia	223 (69.5%)	76 (23.7%)	15 (4.6%)	7 (2.2%)	0 (0%)	321
North America	1837 (44.7%)	1420 (34.5%)	509 (12.4%)	314 (7.6%)	31 (0.8%)	4111
Europe	703 (52.2%)	471 (35.0%)	95 (7.0%)	75 (5.6%)	3 (0.2%)	1347
Africa	100 (80.0%)	21 (16.8%)	2 (1.6%)	2 (1.6%)	0 (0%)	125
South America	544 (74.3%)	136 (18.6%)	26 (3.5%)	19 (2.6%)	7 (1.0%)	732
Oceania	177 (43.8%)	115 (28.5%)	44 (10.9%)	44 (10.9%)	24 (5.9%)	404
Global	3584 (50.9%)	2239 (31.8%)	691 (9.8%)	461 (6.6%)	65 (0.9%)	7040 (100%)

4.3.3 A hydro-climate classification to estimate global flood timing

A rule-based classification (lower panel of Figure 4-7; herein referred to as D10) was developed to partition the land surface into 10 hydro-climate classes (top panel of Figure 4-7). Although it is possible to further break each class into sub-regions and potentially improve the model's predictive power, the classification scheme was kept at this level of simplicity because the tree is found to be able to represent the key regional patterns of the best predictors. Among the 11 proposed separating variables, eight were retained for the final classification (MAP, P_{dry}, P_{swet}, P_{wwet}, $TD_{indicator}$, f_{snow}, T_{hot}, and T_{cold}), which partition the world into five rainfall-dominant classes (Class 1 to Class 5) and five snowmelt-dominant classes (Class 6 to Class 10).

As shown in the resulting tree, the key variable separating rainfall-dominant classes from snowmelt-dominant classes is $TD_{indicator}$, based on

the fact that regions where the transition timing predictor (*TD*) cannot be reliably defined (i.e. $TD_{indicator} = 0$) are unlikely to have snowmelt occurring. For locations where the *TD* predictor can be reliably estimated, the "transitional regions" between rainfall-dominant and snowmelt-dominant groups were identified using the fraction of precipitation falling as snow ($f_{snow} < 0.1$). The key characteristic of these "transitional regions" is a relatively low amount of snowfall (and thus snowmelt) occurring, so that this predictor is unlikely to play a dominant role in flood generation. Across these "transitional regions", rainfall-dominant locations (Class 5) were defined if more than 12% of precipitation falls into the wettest month of summer (i.e. $P_{swet}/MAP > 0.12$), while the other locations were classified as snowmelt-dominant (i.e. Class 6).

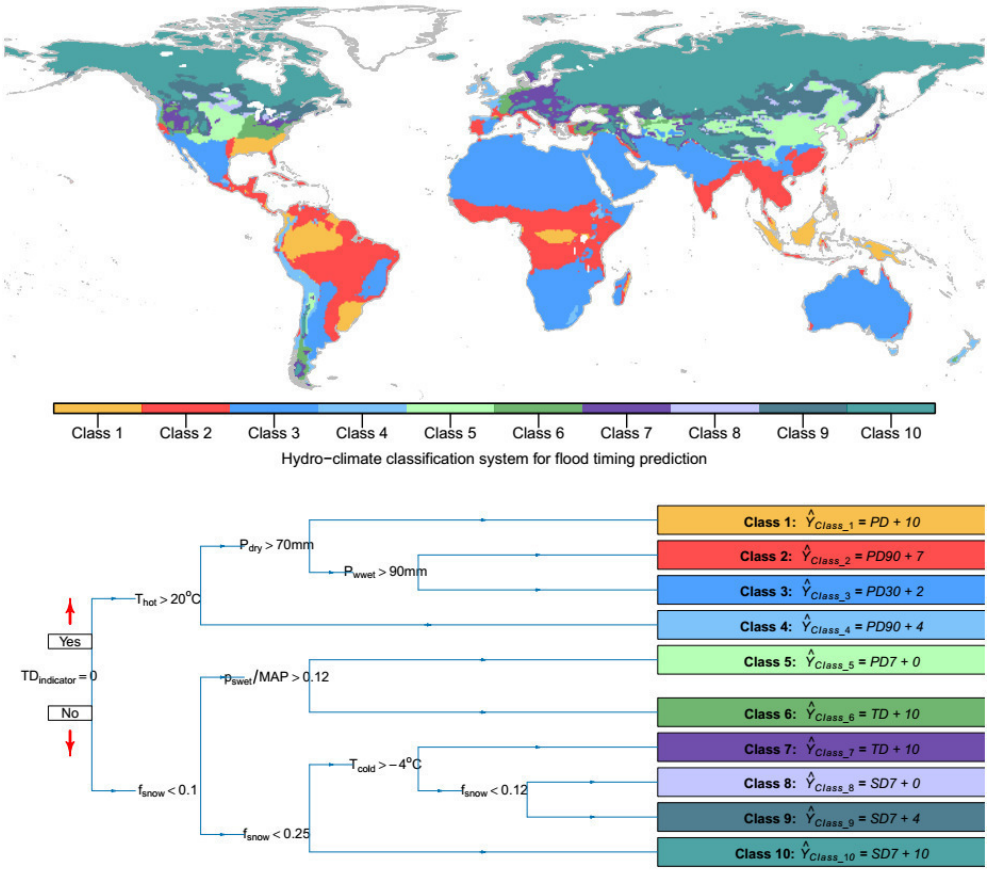


Figure 4-7. Global maps of climate regions (top panel) partitioned by the D10 hydro-climate system (bottom panel). Each hydro-climate class is defined following a set of separation rules and has a prediction of flood timing as a linear function of one predictor.

The subsequent divisions were completed by visually examining the spatial pattern of climate indices (supplementary Figures C1 to C11) to find

the best match to clusters of climate predictors. As summarised in both Figure 4-7 and Table 4-4, temperature of the hottest month (T_{hot}), the amount of precipitation that falls within the driest month (P_{dry}) and the wettest month in winter (P_{wwet}) were selected as separating variables for rainfall-dominant flood classes, while the fraction of precipitation falling as snow (f_{snow}) and the temperature of the coldest month (T_{cold}) are important to derive the boundaries among snowmelt-dominant classes.

The dominant atmospheric predictor of flood timing was then identified for each hydro-climate class to form a linear function between that predictor and the flood timing response. The most relevant predictor and associated lag-day in each class were identified through the optimisation process described in Section 4.2.3.2 and are presented in Table 4-4. Among 7,040 locations, the prediction scheme could be applied for 6,757 stations in total (excluding 283 stations due to a missing value of the identified predictor). The majority of 'no prediction' locations falls into Class 1 due to the non-seasonal characteristic of rainfall-predictors across the south-eastern U.S., which contains most stations classified into Class 1. The maximum value of γ_{R_j} across ten hydro-climate classes was found to be 10 days, indicating that floods, on average, occur within the 10-day window after the timing of the dominant predictor. Prediction errors (represented by mean absolute error) ranges from 19.9 days (Class 3) to 48.7 days (Class 7), and when averaged across all stations had a value of 31.9 days. Across all land regions, flood timing is mainly predicted by snowmelt and long-term precipitation predictors, which respectively explain flood timing for 48% and 42% of the global landmass.

Table 4-4. Description of the hydro-climate classes defined through the D10 classification system (lower panel of Figure 3-6). 'No prediction' indicates locations where there is no predictor available to predict flood timing.

Class	Climate indices used to define hydro-climate class	Number of gauges	% of no prediction	Dominant flood generation	Lag-day	Prediction errors (MAE; in days)	% of global land mass
1	$TD_{indicator}$, T_{hot}, P_{dry}	502	33	Short-term precipitation (PD)	10	35.5	4.6
2	$TD_{indicator}$, T_{hot}, P_{dry}, P_{wwet}	902	4	Long-term precipitation (PD90)	7	23.4	14.6
3	$TD_{indicator}$, T_{hot}, P_{dry}, P_{wwet}	709	4	Long-term precipitation (PD30)	2	19.9	25.1
4	$TD_{indicator}$, T_{hot}	672	5	Long-term precipitation (PD90)	4	26.4	2.3
5	$TD_{indicator}$, f_{snow}, P_{swet}, MAP	396	4	Short-term precipitation (PD7)	0	39.0	4.9
6	$TD_{indicator}$, f_{snow}, P_{swet}, MAP	1,082	0	Snowmelt predictor (TD)	10	33.7	2.3
7	$TD_{indicator}$, f_{snow}, T_{cold}	905	0	Snowmelt predictor (TD)	10	48.7	3.0
8	$TD_{indicator}$, f_{snow}, T_{cold}	233	0	Snowmelt predictor (SD7)	0	18.1	1.3
9	$TD_{indicator}$, f_{snow}, T_{cold}	775	0	Snowmelt predictor (SD7)	4	31.7	6.8
10	$TD_{indicator}$, f_{snow}	864	1	Snowmelt predictor (SD7)	10	34.0	35.1
Global	$TD_{indicator}$, T_{hot}, P_{dry}, P_{wwet}, P_{swet}, MAP, f_{snow}, and T_{cold}	7,040	4	-	-	31.9	100.0

Although the overall performance of the prediction scheme is reasonable at the global scale, there are some regions that have a large prediction error (Figure 4-8) such as central North America or the Alps. The characteristics of flood generation across these regions suggests two main reasons for poor performance: (1) the prediction scheme does not correctly define the most important flood generation mechanism over some relatively small geography areas with large variability of identified predictor (partly due to the coarse resolution of climate reanalysis products), and/or (2) there is more than one mechanism that contributed substantially to flood generation (e.g. extreme rainfall events in summer floods and snowmelt events in winter-spring floods) and using a single most important predictor may not capture this complexity. The latter reason is generally consistent with the results of the seasonality assessment for flood timing (reported in Section 4.3.1), which did not reject the null-hypothesis of no seasonality for many stations over central North America or the Alps.

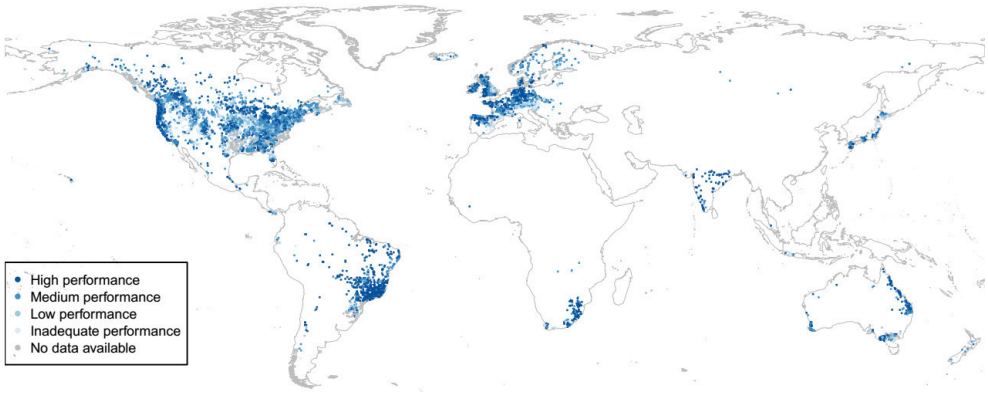


Figure 4-8. Prediction errors across 7,040 stations grouped into the five consistency definitions in Table 4-1 based on local performance.

The global prediction of flood timing using the proposed classification system (Figure 4-9), however, can reflect most of the large-scale spatial association in flood timing, especially in the southern hemisphere, where rainfall plays the key role in flood generation. The longitudinal transition over regions with high station density (e.g. North America and Europe) is also generally illustrated, suggesting the potential capacity of this prediction scheme in representing the spatial complexity of flood generation processes. The prediction of flood timing not only has consistency with flood timing based on regional observational studies in Europe and North America, but

also has high consistency with the spatial patterns of the main high-flow season obtained from the global hydrological model (Lee et al., 2015). Additionally, the predicted flood timing compares favourably to the streamflow peak month identified from an investigation using monthly stream flow series across 1345 sites globally (Dettinger and Diaz, 2000), providing confidence that a relatively simple predictive scheme—based on readily available atmospheric predictors obtained from reanalysis datasets—is able to provide credible predictions of flood timing in both data rich and sparse regions.

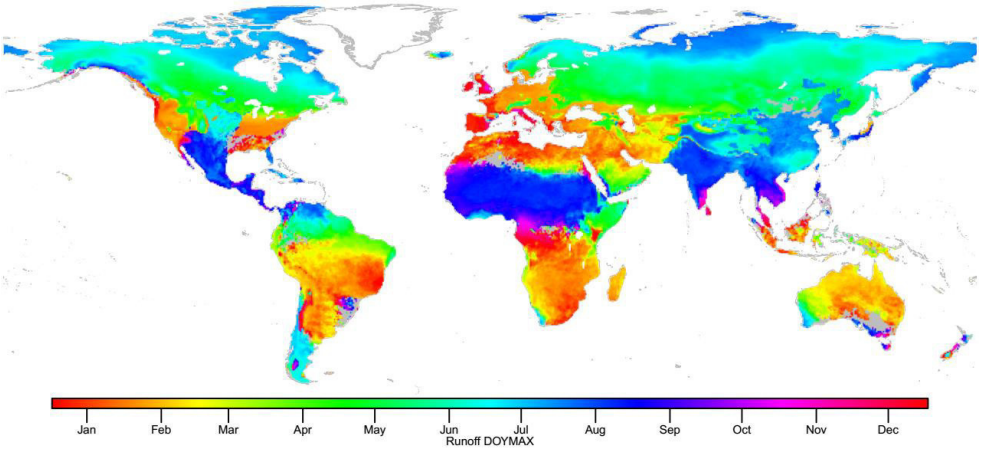


Figure 4-9. Global prediction of flood timing using reanalysis climate forcing datasets and D10 decision tree. Grey colour indicates locations where there is no suitable predictor available due to lack of seasonality.

4.4 Summary and conclusions

This study analysed the spatial consistency of observed flood seasonality from 7,894 streamflow records (Do et al., 2018b, Gudmundsson et al., 2018b) and climate variables derived from an atmospheric forcing reanalysis dataset (Dee et al., 2011). The analysis has not only demonstrated consistent results with existing studies of flood seasonality across Europe and North America (Villarini, 2016, Hall and Blöschl, 2018, Blöschl et al., 2017, Burn and Whitfield, 2016), but has facilitated the extension of flood timing estimates across the globe. Having identified spatial consistency between flood timing and selected variables representing flood generating mechanisms, this study provides important observation-based evidence of homogeneous regions of flood generation mechanisms. Short-term precipitation predictors are the dominant driver of flood timing in the south-eastern region of the US,

northern Australia, and the southern and eastern regions of Brazil; long-term precipitation predictors are the dominant driver in central Brazil, western Europe, and southern Australia; and snowmelt predictors are the dominant driver in the high-latitude areas of the North American and Eurasian continents. These findings complement current understanding of flood seasonality characteristics at the global scale.

Notwithstanding the complexity of dominant flood producing mechanisms, this study was able to empirically identify a low discrepancy between flood timing and a single most important atmospheric predictor, yielding a high percentage of locations with discrepancy of less than or equal to 45 days, i.e. flood timing and the most suitable predictor occur in the same season (continental scale: 73% – 94%, global average 82%). Taking advantage of the strong agreement between flood timing and climate predictors, a rule-based classification system was developed to partition the world into ten hydro-climate classes, each representing regions sharing a common flood generation mechanism. The classification was used to infer flood timing globally, including regions not covered by streamflow gauges. Although there are some regions with a high prediction error (e.g. central North America, the Alps and southern Australia), the proposed model, which has a relatively simple structure, performs well in predicting flood timing (global mean absolute error of 31.9 days) and was able to preserve large-scale spatial associations in flood timing across the globe. The spatial pattern of flood seasons obtained from this analysis compares favourably to the high-flow seasonal data obtained from a global hydrological model (Lee et al., 2015) and streamflow peak month obtained from 1345 sites globally (Dettinger and Diaz, 2000).

The classification system proposed in this study can be used to define regions of similar flood generating processes at the global scale. Considering its relative simplicity and reproducible character, the present rule-based classification tree could also be used as a common classification system for different climate datasets to assess the variation in either flood timing or flood-generating processes. For example, the classification tree could potentially be used to assess the effect of projected future climates in terms

of the timing of floods. Finally, this classification could be used as a measure of global hydrological model performance, by providing an indicator that these models correctly simulate the climatic mechanisms that lead to large streamflow events.

Data and Acknowledgments

Observational streamflow index are taken from the GSIM archive – freely available from http://dx.doi.org/10.1594/PANGAEA.887470 (Gudmundsson et al., 2018a) and http://dx.doi.org/10.1594/PANGAEA.887477 (Do et al., 2018a). Gridded precipitation and temperature data are taken from ERA-Interim global atmospheric reanalysis and are available at https://www.ecmwf.int (Dee et al., 2011). Mr Hong Xuan Do receives financial support from the Australia Award Scholarship (AAS) and D R Stranks Travelling Fellowship. This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

Supplementary materials for Chapter 4

1 Methodologies

1.1 Circular statistics

At each station, the ordinal day of the annual maximum streamflow was determined for each year in the 1981-2010 reference period and the average day of flood timing was determined using circular statistics (Mardia and Jupp, 2009).

The ordinal day, D_i, of annual maximum streamflow in year i is converted into an angular value θ_i (in radians):

$$\theta_i = D_i \frac{2\pi}{m_i} \qquad 0 \le \theta_i \le 2\pi$$

where $D_i = 1$ corresponds to January 1, and $D_i = m_i$ corresponds to December 31, and where m_i is the number of days in that year (i.e. 365 or 366 days). The average date of occurrence of flood, $\overline{D}$, at a station is defined by

$$\overline{D} = \begin{cases} tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right)\left(\frac{\overline{m}}{2\pi}\right) & \bar{x} > 0, \bar{y} \ge 0 \\ tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right)\left(\frac{\overline{m}}{2\pi}\right) + \pi & \bar{x} \le 0 \\ tan^{-1}\left(\frac{\bar{y}}{\bar{x}}\right)\left(\frac{\overline{m}}{2\pi}\right) + 2\pi & \bar{x} > 0, \bar{y} < 0 \end{cases}$$

with $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\cos(\theta_i)$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\sin(\theta_i)$ are respectively the cosine and sine components of the average date, and , $\overline{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$ is the average number of days per year.

The concentration R (ranges from 0 to 1) of DOYMAX represents the strength of the seasonality in flood timing, and is defined by:

$$R = \sqrt{\bar{x}^2 + \bar{y}^2} \qquad 0 \le R \le 1$$

A value of R = 0 indicates that flood occurrence dates were spread throughout the year, while R = 1 indicates that all flooding events at a specific location occur on the same ordinal day across all years. A circular version of Kuiper's test (Mardia and Jupp, 2009) was conducted to check whether the time series is circularly uniform, and only stations that reject the null

hypothesis of circular uniformity at the 10% confidence level (7,040 stations) were considered for the development of flood timing prediction.

The same technique was also applied to climate indices derived from atmospheric reanalysis to generate records of predictors representing flood-producing mechanisms. For each predictor of flood timing, missing value was assigned to grid cells where the null hypothesis of circular uniformity cannot be rejected at the 10% confidence level.

1.2 Degree-day method to simulate snowmelt dynamic

A simple degree-day model (Woods, 2009, Hock, 2003) was used to simulate snowmelt dynamic and derive two predictors of flood timing: date of the peak value of daily snowmelt or rain-on-snow (SD) and Date of the peak value of 7-day snowmelt or rain-on-snow (SD7). The model is represented through below equations:

$For\ T \leq T_{crit}$

$$S_{snow}(t) = S_{snow}(t-1) + P(t)$$
$$P_{snow}(t) = 0$$

$For\ T > T_{crit}$

$$P_{snow}(t) = \min\left(f_{dd} * \max(T(t) - T_{crit}, 0), S_{snow}(t-1)\right) + P(t)$$
$$S_{snow}(t) = \max(S_{snow}(t-1) - P_{snow}(t), 0)$$

where T (°C) is the air temperature, T_{crit} is the critical temperature for the degree day method (set at 0°C to be consistent with the calculation of TD predictor), S_{snow} is the daily snow storage (mm/day), which is assumed to melt when air temperature T is above critical value. P is the daily precipitation (mm/day) and is assumed to fall as snow (and contribute to the snowpack) when $T < T_{crit}$, and as rainfall (and contribute to rain-on-snow event) when $T > T_{crit}$. Variable P_{snow} is the combination of snowmelt and rainfall (mm/day), and f_{dd} is the melt rate, which was set at 3.0 (mm/d/C), as previously used in Woods (2009).

Having calculated a time series of P_{snow} for each day of the year, the predictor SD is defined as the ordinal date of the maximum value of P_{snow}. This variable, in the absence of significant amount of snowfall, is simply a

measure of precipitation and thus the fraction of total precipitation falling as snow (f_{snow}) is used to assign missing data for this predictor in locations with less than 10% of total precipitation falls as snow.

2 Supplementary figures

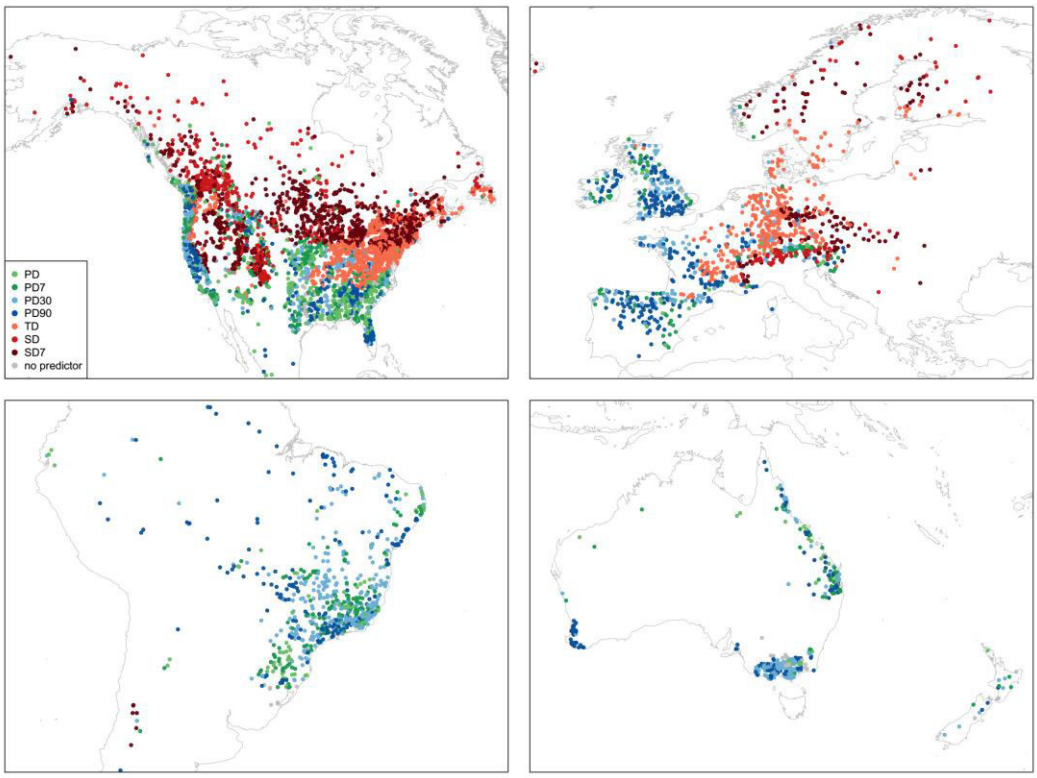


Figure S4-1. Maps of the predictor with the smallest discrepancy to flood timing in regions with a high density of stations (for 7,040 locations with seasonal flood timing). Top left: North America; top right: Europe; bottom left: South America; and bottom right: Australia and New Zealand.

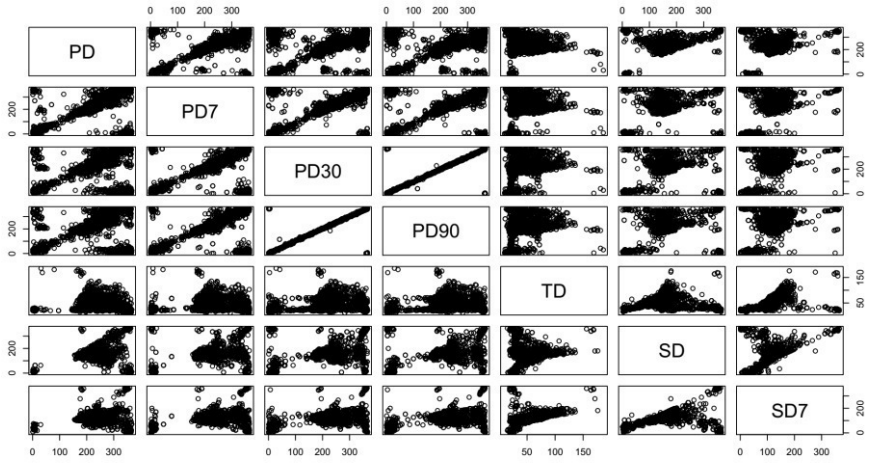


Figure S4-2. Scatter plots for the seven predictors across 7,040 locations with seasonal flood timing. There is a high correlation between short-precipitation predictors (PD and PD7) and

long-precipitation predictors (PD30 and PD90), indicating potential noise in determining the most important predictor.

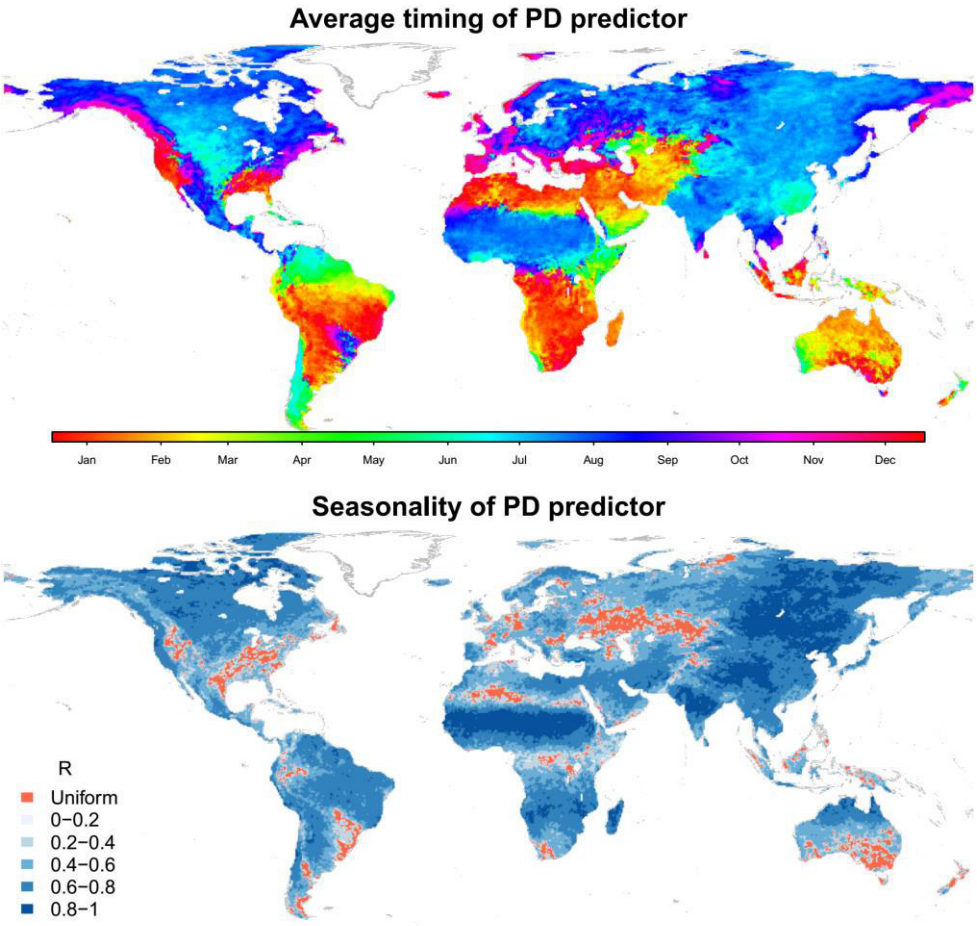


Figure P4-1. Seasonality of PD predictor derived from ERA-Interim dataset for the period 1981-2010. Top panel: long-term-mean value of the predictor. Lower panel: concentration index (R) of the predictor (values range from 0 to 1); red cells represent locations where circular hypothesis was not rejected at the 10% significance.

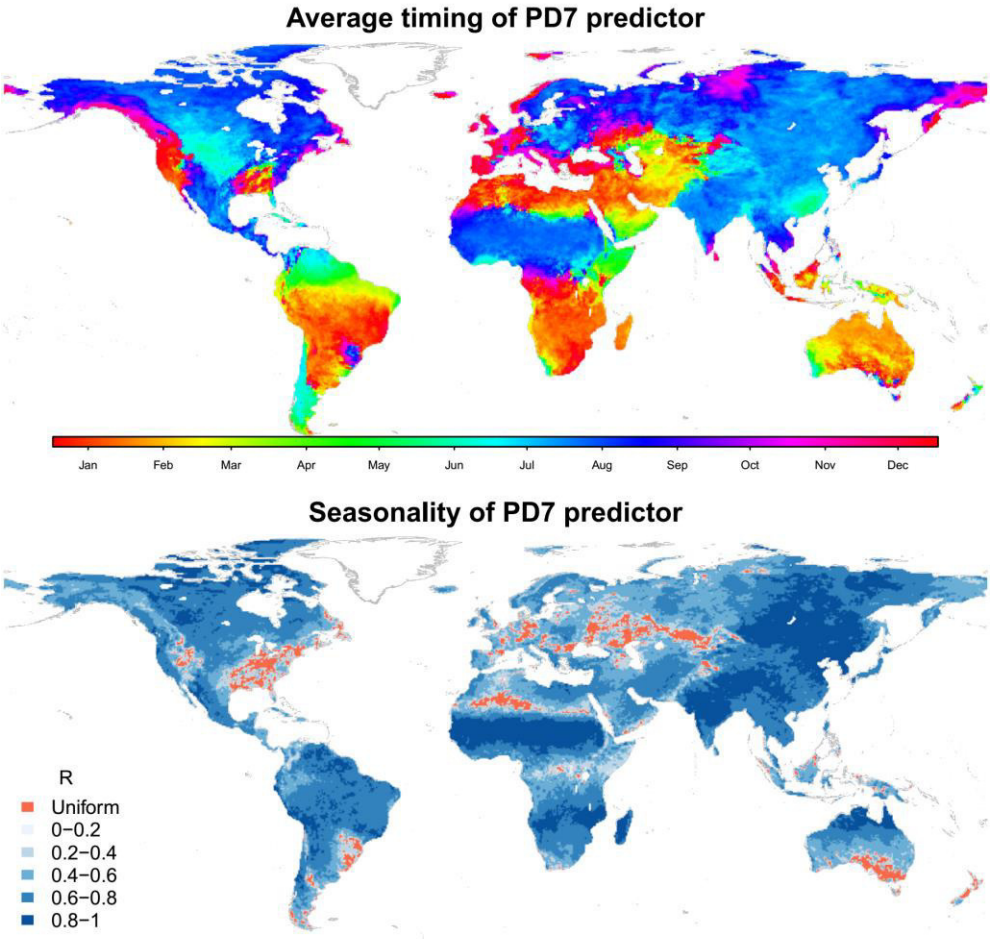


Figure P4-2. Similar to Figure P4-1, but for PD7 predictor.

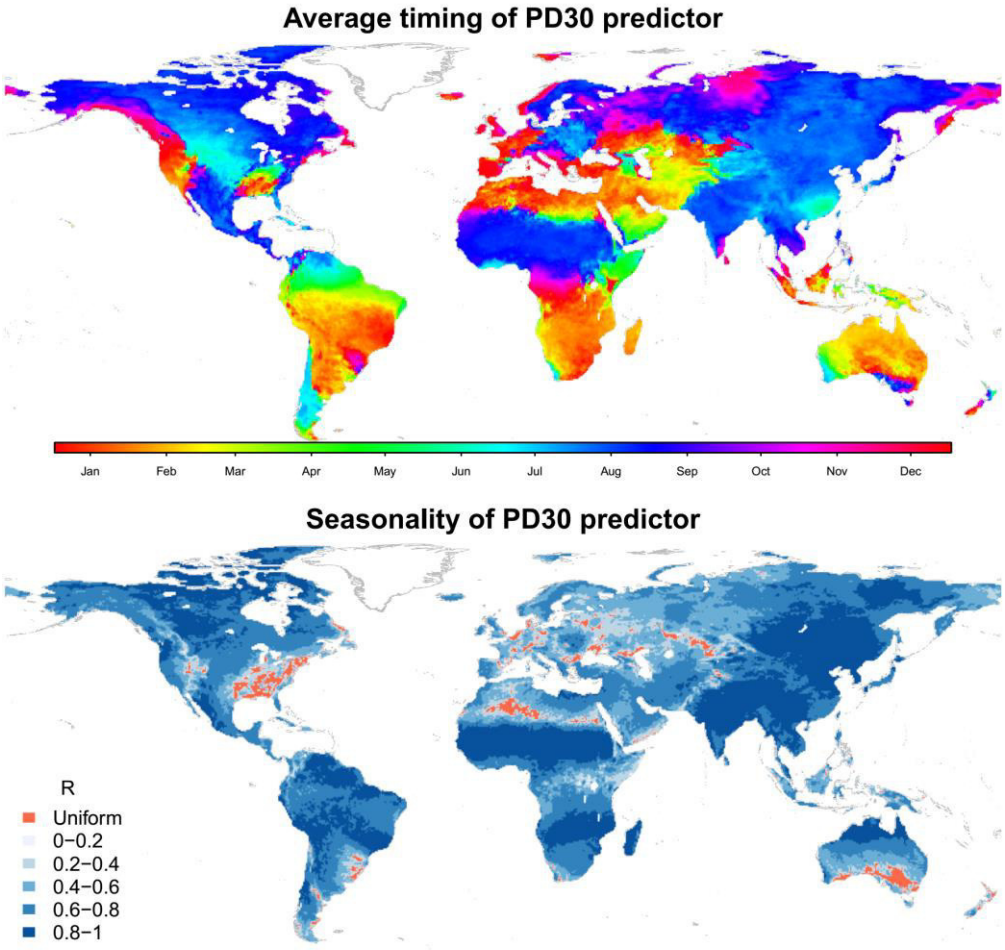


Figure P4-3. Similar to Figure P4-1, but for PD30 predictor.

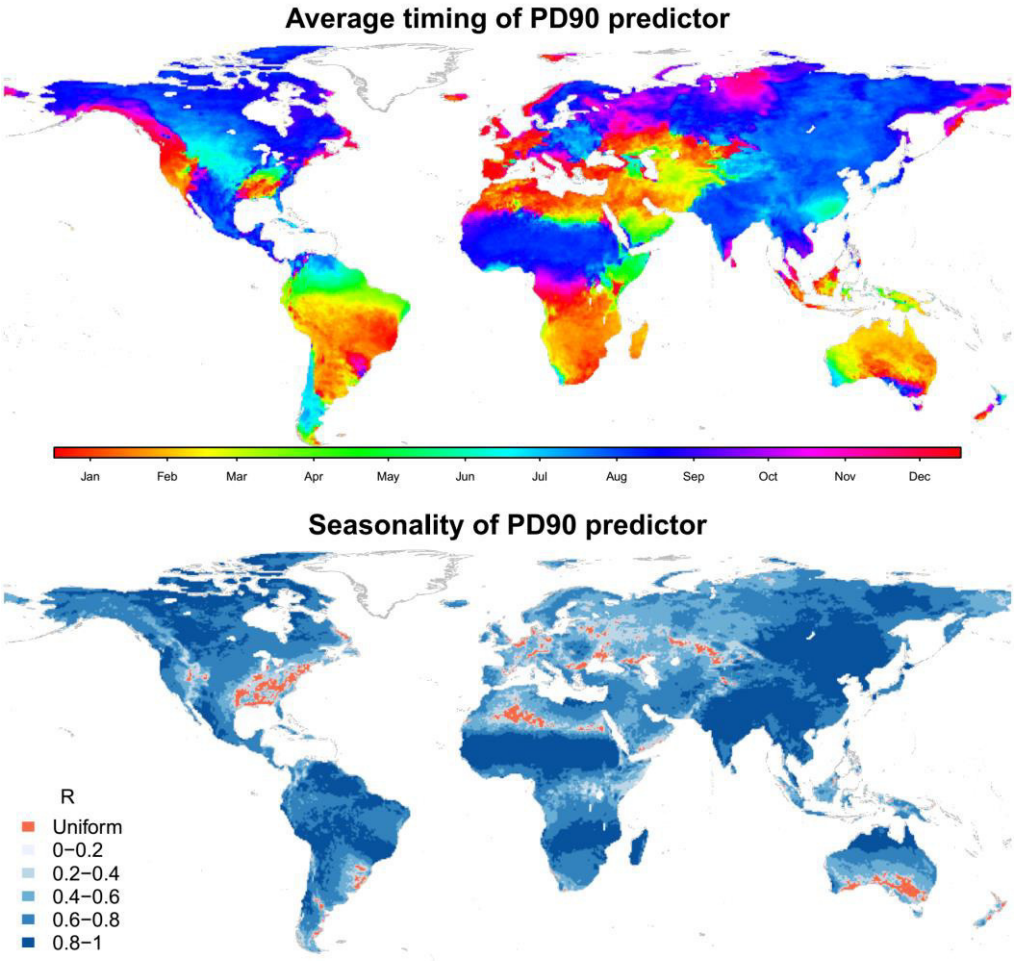


Figure P4-4. Similar to Figure P4-1, but for PD90 predictor.

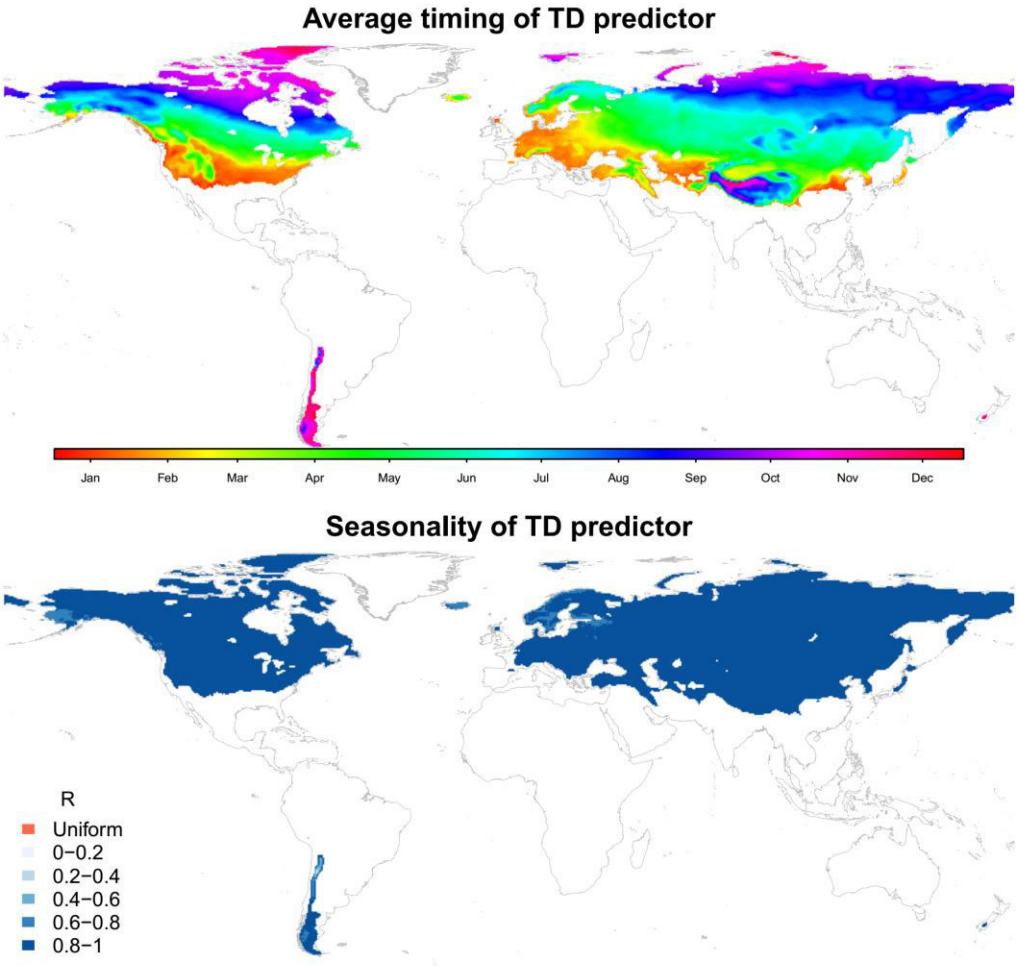


Figure P4-5. Similar to Figure P4-1, but for TD predictor. Here the white spaces represent locations where TD cannot be reliably defined for at least 70% of the years (note that Greenland and Antarctica were excluded from this analysis).

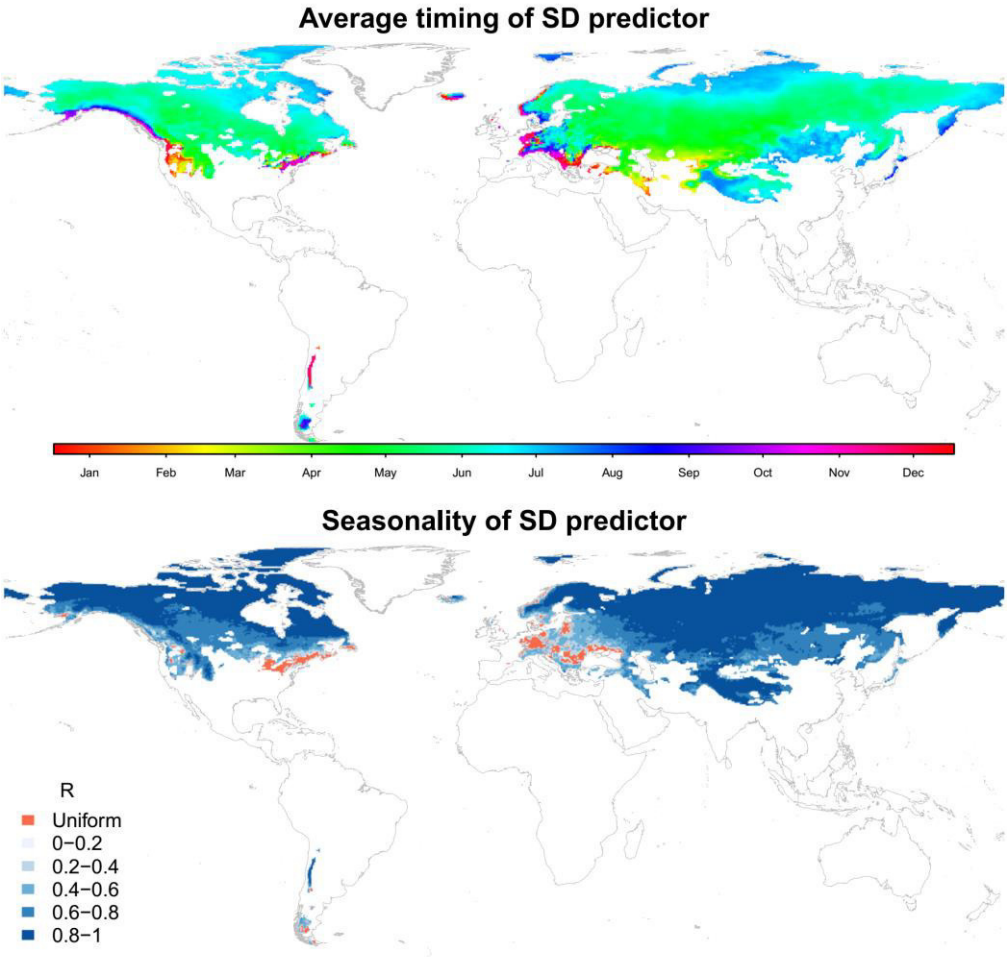


Figure P4-6. Similar to Figure P4-1, but for SD predictor. Here the white spaces represent locations with less than 10% of precipitation fall as snow (note that Greenland and Antarctica were excluded from this analysis).

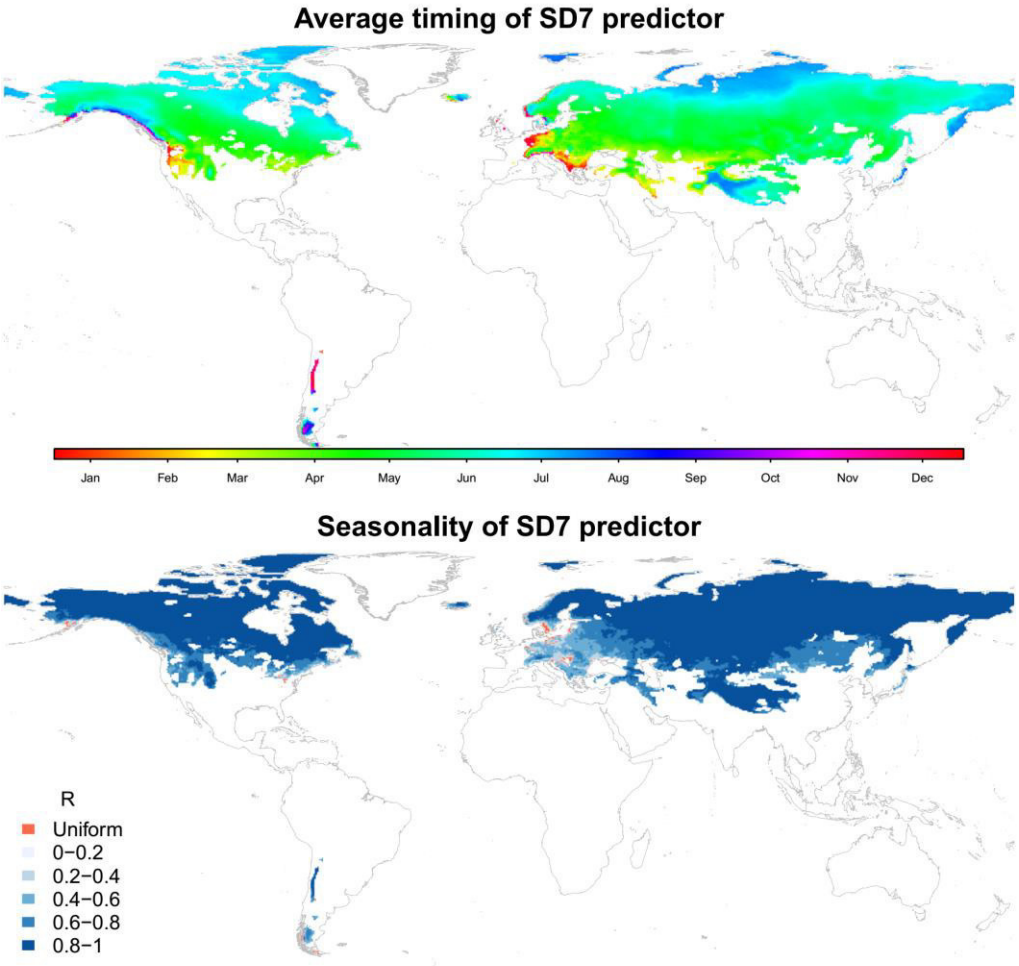


Figure P4-7. Similar to Figure P4-6, but for SD7 predictor.

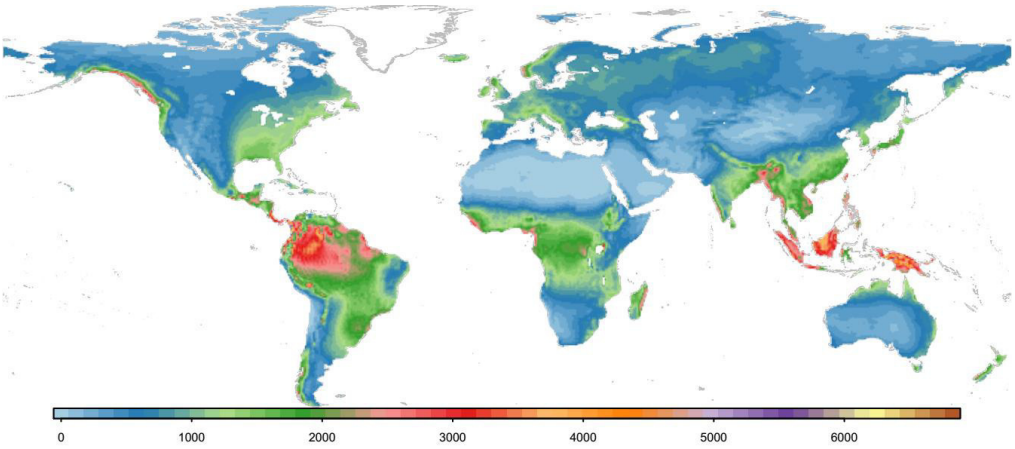


Figure C4-1. Spatial variation of splitting variable MAP (mean annual precipitation; in mm)

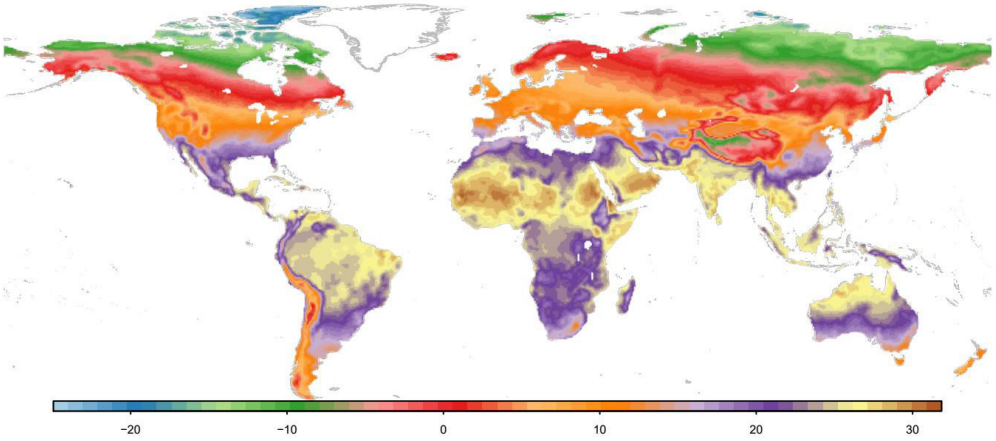


Figure C4-2. Spatial variation of splitting variable MAT (mean annual temperature; in oC)

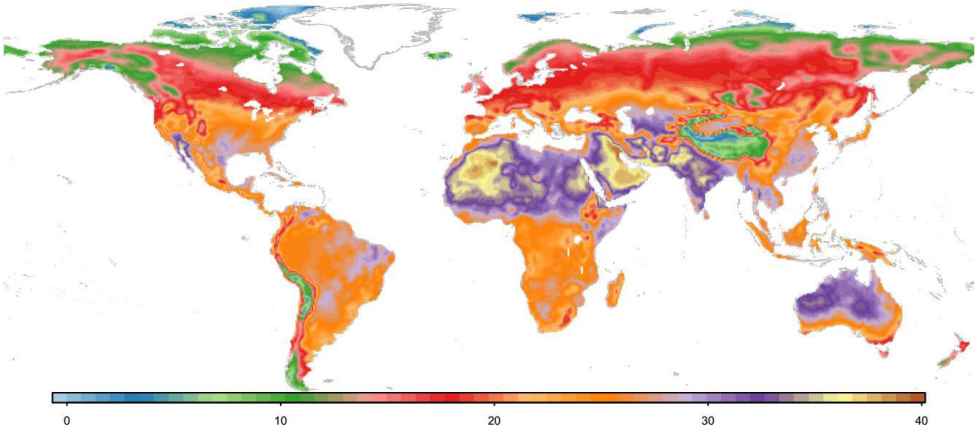


Figure C4-3. Spatial variation of splitting variable Thot (temperature of the hottest month; in oC)

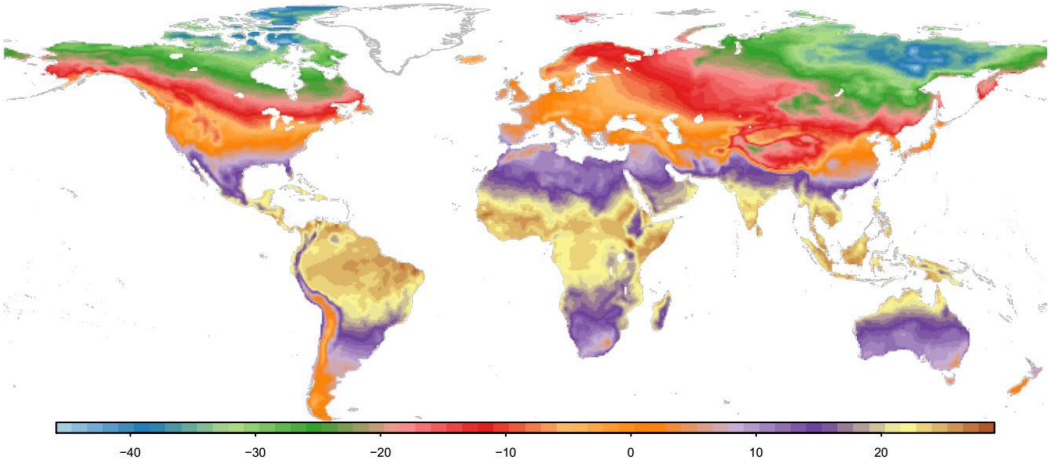


Figure C4-4. Spatial variation of splitting variable Tcold (temperature of the coldest month; in oC)

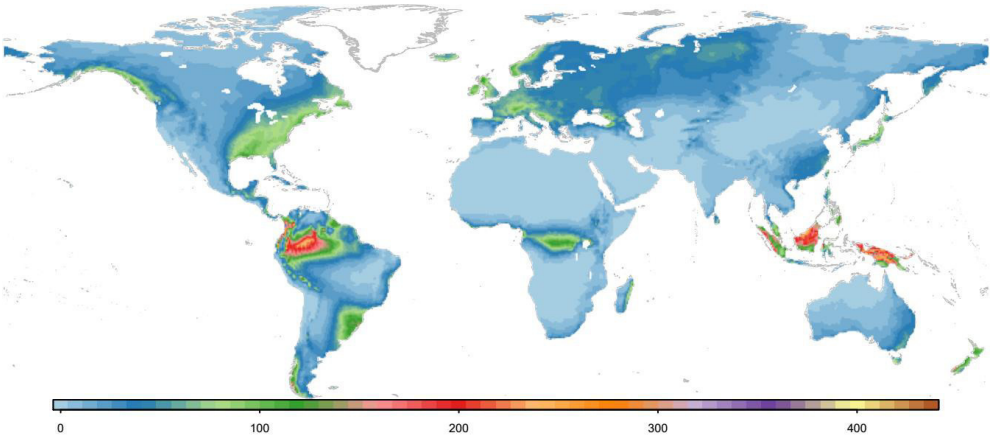


Figure C4-5. Spatial variation of splitting variable Pdry (precipitation of the driest month; in mm)

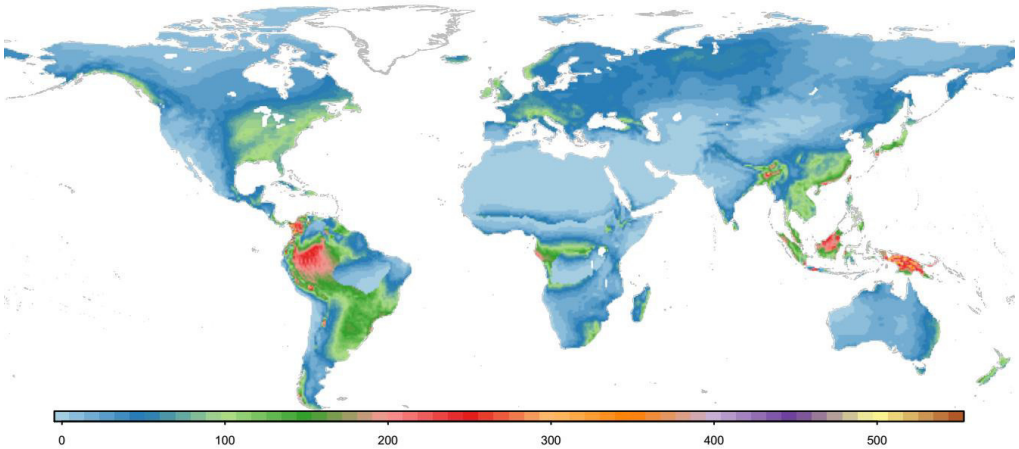


Figure C4-6. Spatial variation of splitting variable Psdry (precipitation of the driest month in summer; in mm). Summer (winter) is defined as the warmer (cooler) six month period of October – March and April – September for each respective hemisphere.

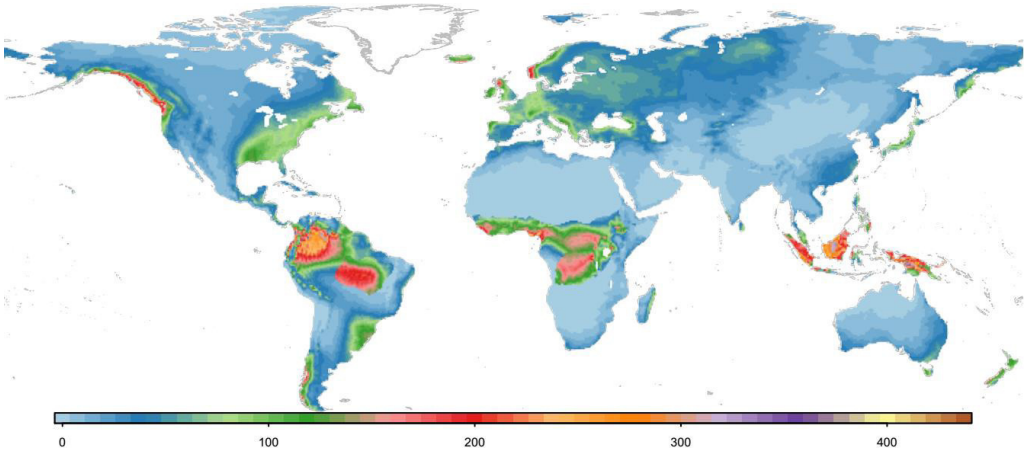


Figure C4-7. Spatial variation of splitting variable Pwdry (precipitation of the driest month in winter; in mm).

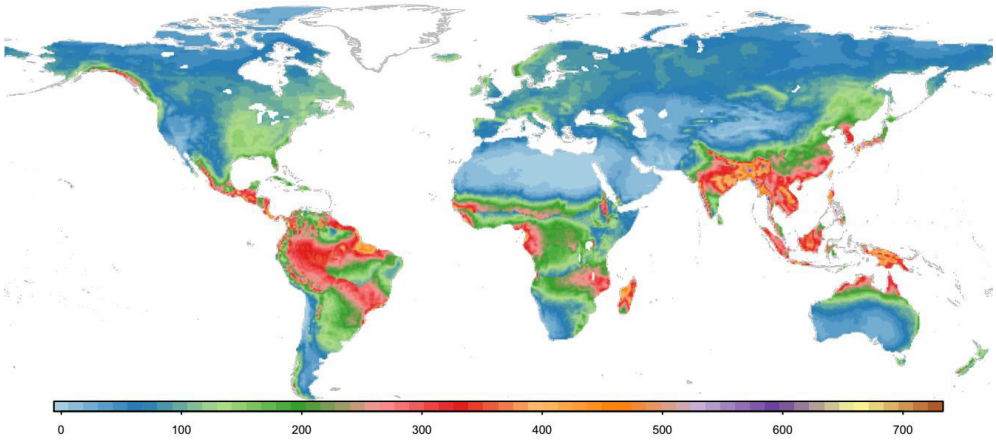


Figure C4-8. Spatial variation of splitting variable Pswet (precipitation of the wettest month in summer; in mm).

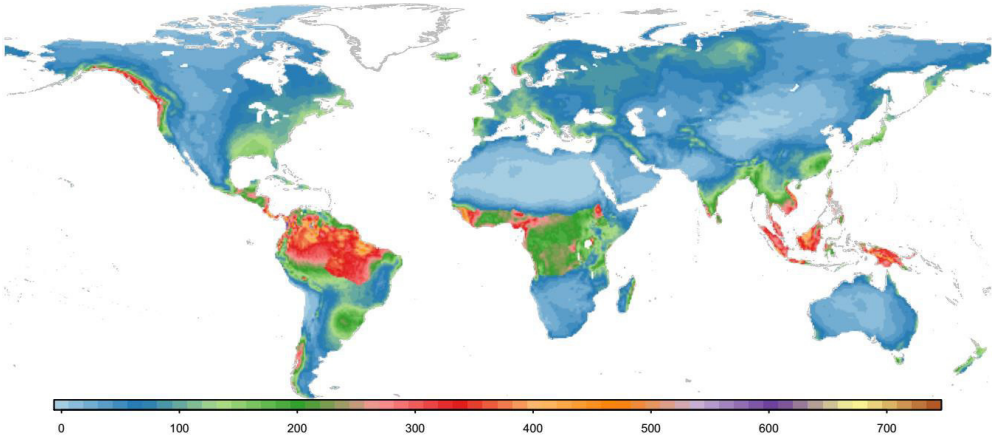


Figure C4-9. Spatial variation of splitting variable Pwwet (precipitation of the wettest month in winter; in mm).

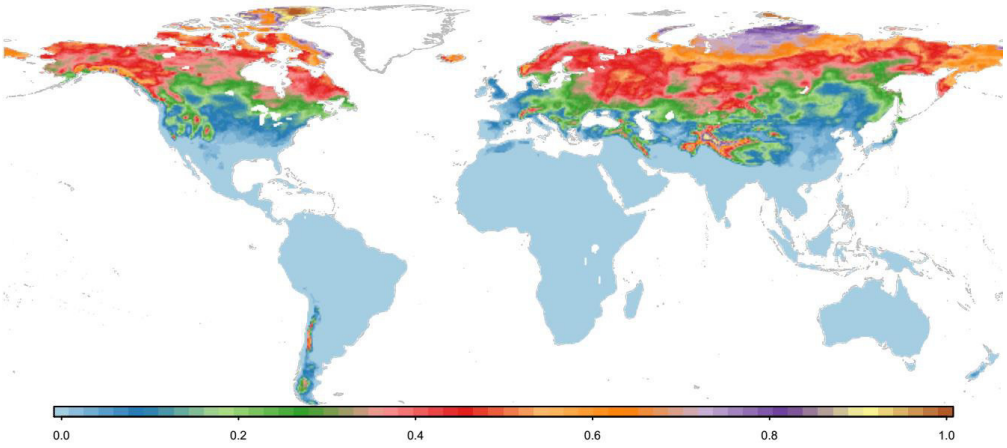


Figure C4-10. Spatial variation of splitting variable fsnow (fraction of precipitation falls as snow; precipitation is defined as snowfall for days with temperature below 0oC).

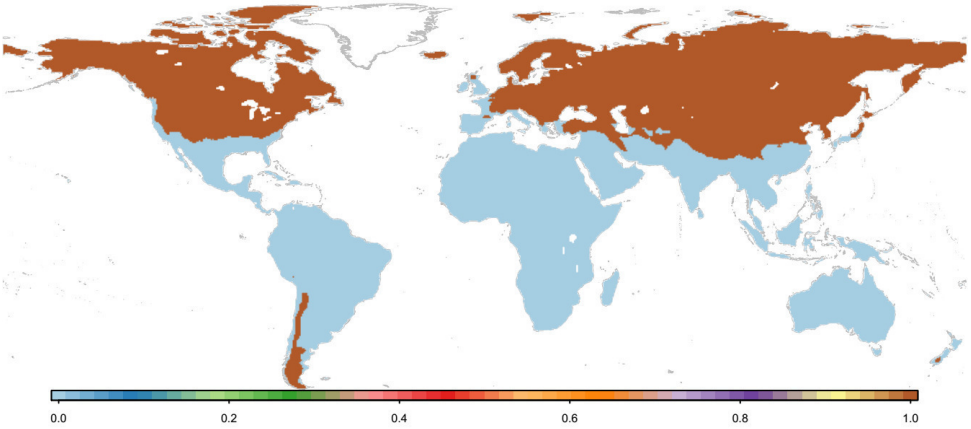


Figure C4-11. Spatial variation of splitting variable TDindicator (binary variable where a value of 0 indicates predictor TD cannot be defined for at least 70% of the years).

Chapter 5. Historical and future changes in global flood magnitude – Evidence from a model-observation investigation (Paper 4)

Statement of Authorship

Title of Paper	Historical and future changes in global flood magnitude – Evidence from a model-observation investigation
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	N/A

Principal Author

Name of Principal Author (Candidate)	Hong Xuan Do		
Contribution to the Paper	Designed the experiments, conducted statistical analyses, analysed results, wrote manuscript and acted as corresponding author.		
Overall percentage (%)	80		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	07/02/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Fang Zhao		
Contribution to the Paper	Suggest scope of study and refined research questions, helped to reprocess data, design experiments and interpret results, and provided feedbacks on manuscripts.		
Signature		Date	13/02/19

Name of Co-Author	Seth Westra		
Contribution to the Paper	Suggest scope of study, helped to design experiments and interpret results, provided feedbacks on manuscripts.		
Signature		Date	15/02/19

Name of Co-Author	Michael Leonard		
Contribution to the Paper	Suggest scope of study, helped to design experiments and interpret results, provided feedbacks on manuscripts.		
Signature		Date	19/02/2019

Name of Co-Author	Lukas Gudmundsson		
Contribution to the Paper	Suggest scope of study, helped to design experiments and interpret results, provided feedbacks on manuscripts.		
Signature		Date	11/02/2019

Abstract

To improve understanding of trends in extreme flows related to flood events at the global scale, historical and future changes of annual maximum streamflow are investigated, using a comprehensive streamflow archive and six global hydrological models. The models' capacity to characterise trends in annual maximum streamflow is evaluated across 3666 locations (1971-2005 period), focusing on four aspects of trends: (i) mean, (ii) spread, (iii) percentage of locations showing significant trends and (iv) spatial pattern. Compared to observed trends, simulated trends over the historical period generally have higher mean, lower spread, a similar percentage of locations showing significant (increasing/decreasing) trends and low-to-moderate pattern similarity (driven by atmospheric forcing). For projected simulations (2006-2099 simulation period), simulated trends have relatively low spread in the trend mean (ranging from -1.1% to 1.0% change per decade) and trend standard deviation (ranging from 1.8% to 4.4% change per decade). However, the ensembles have high uncertainty in the spatial structure of the trends, with low inter-model correlation (from -0.17 to 0.23). Under a 'business-as-usual' socio-economic scenario, most discharge simulations project an increasing trend in flood magnitude over Siberia, Greenland, South and South-East Asia while southern Australia, the Mediterranean, and eastern Europe are projected with a decreasing trend. High-risk regions (i.e. projected with an increasing trend by most models) are sparsely sampled, represented by less than 1% of all gauges available in the to-date most comprehensive stream-gauge catalogue, indicating a highly uncertain future for flood-prone communities in the context of climate change.

5.1 Introduction

Global hydrological models (GHMs) are critical tools for diagnosing factors of rising trends in flood risk (Munich Re, 2015, Swiss Re, 2015, Miao, 2018, Smith, 2003, Guha-Sapir et al., 2015, CRED, 2015), and can help identify the contribution of changing flood hazard characteristics relative to the changing exposure of human assets to floods. GHMs are also used to project future changes in flood hazard, owing to their ability to simulate streamflow under projected atmospheric forcing. Using GHM simulations, several studies have found more regions showing increasing trends than decreasing trends in flood hazards at the global scale, and have attributed these changes to anthropogenic climate change (Dankers et al., 2014, Arnell and Gosling, 2014, Alfieri et al., 2015, Kettner et al., 2018). The pattern of increasing trends obtained from GHM simulations is consistent with observations of increases in precipitation extremes (Westra et al., 2013, Westra et al., 2014, Donat et al., 2013, Guerreiro et al., 2018) that have been used by a number of studies as a proxy to suggest that flood hazard may increase as a result of climate change (Alfieri et al., 2017, Pall et al., 2011, IPCC, 2012, Forzieri et al., 2016).

The inference of changes in flood hazard following the same direction as extreme precipitation may be appropriate over specific regions (Hoegh-Guldberg et al., 2018, Mallakpour and Villarini, 2015, Mangini et al., 2018), but is potentially inapplicable globally. This is due to a ‘dichotomous relationship’ between trends exhibited in extreme precipitation and extreme streamflow (Sharma et al., 2018), highlighted in recent observation-based studies of streamflow magnitudes (Wasko and Sharma, 2017, Do et al., 2017, Hodgkins et al., 2017, Gudmundsson et al., 2019). The hypothesised reason for this potentially inconsistent relationship is the complexity of the drivers of flood risk (Johnson et al., 2016, Blöschl et al., 2017, Berghuijs et al., 2016), with the implication that historical and future changes to flood hazard at the global scale are unlikely to be reflected by changes to a single proxy variable such as annual maximum rainfall in isolation. For example, even though trends in extreme flows are highly correlated to changes in extreme rainfall when rainfall plays the dominant role (Mallakpour and Villarini, 2015,

Blöschl et al., 2017), snowmelt-related flood magnitude has been found to decrease in a warmer climate, potentially due to a shift in snowmelt timing (Burn and Whitfield, 2016, Cunderlik and Ouarda, 2009). The sign of change can remain unclear for locations where antecedent soil moisture plays an important role (Woldemeskel and Sharma, 2016, Sharma et al., 2018), owing to the influences of seasonal/annual precipitation, potential evaporation and extreme precipitation (Bennett et al., 2018, Ivancic and Shaw, 2015, Leonard et al., 2008).

To better understand trends in streamflow extremes, the emphasis has therefore moved to analysing trends in direct streamflow measurements. Investigations using streamflow observations at global, continental and regional scales (see Do et al. (2017) and references therein) have generally detected a mixed pattern of trends, with some global-scale studies finding more stations having decreasing trends than increasing trends (Do et al., 2017, Hodgkins et al., 2017, Kundzewicz et al., 2004). These conclusions appear *prima facie* to be inconsistent with model-based evidence, which generally suggests the opposite (more locations showing increasing trends). However, varying sampling strategies, statistical techniques and reference periods make it difficult to derive a common perspective of trends in global flood hazards from a composite of observational and modelling studies. In addition, data coverage limitations (Hannah et al., 2011, Gupta et al., 2014, Do et al., 2018a) remain a barrier to reliably benchmarking trends over some areas, such as the flood-prone regions of South and East Asia.

GHMs provide an alternative line of evidence about historical and future trends that are able to provide better spatial coverage, and can also enable 'counterfactual' experiments to explore the role of atmospheric forcing, land use change and other drivers of change on streamflow trends. However, unlike climate models, for which the performance in terms of reproducing trends of extreme precipitation has been evaluated substantially (Kiktev et al., 2003, Kiktev et al., 2007, Kumar et al., 2013, Sakaguchi et al., 2012), the performance of GHMs has been assessed mostly on their capacity to represent physical features of the hydrological regime, such as streamflow percentiles, the seasonal cycle or the timing of peak discharge (Gudmundsson et al.,

2012a, Zaherpour et al., 2018, Beck et al., 2017, Zhao et al., 2017, Veldkamp et al., 2018, Pokhrel et al., 2012, Biemans et al., 2011). Streamflow variability, however, can be subject not only to long-term changes in atmospheric forcing, but also to climate variability (e.g. inter-annual, inter-decadal) as well as human activities across the drainage basin (Zhang et al., 2015, Zhan et al., 2012). Thus the capacity to represent physical features of a hydrological model is not necessarily sufficient to determine the model's performance in simulating characteristics of trends in extremes.

To better understand the capacity of GHMs in simulating historical trends in extreme streamflow and potential implications for the development of projections, this study focusses on three research objectives. The first objective is to evaluate the capacity of GHMs (available through the Inter-Sectoral Impact Model Intercomparison Project) to simulate trends in observed streamflow extremes. The particular interest is in reconciling observed and simulated trends in historical streamflow extremes at the global and continental scale using the largest possible streamflow observations database (the Global Streamflow Indices and Metadata archive). The second objective is to determine the representativeness of observation locations in GHM simulations by comparing trends at observation locations to trends across all available land areas. This objective is motivated by the sparse coverage of streamflow observations over several regions, which could lead to biased inferences over large spatial domains wherever gauges are not representative of that domain. The third and final objective is to assess the implication of model uncertainty for projections of flood hazard, focusing on the uncertainty of the mean/spread of trends together with the spatial pattern of trends in annual maximum streamflow.

The rest of this paper is organised as follows. Section 5.2 describes the observed and simulated streamflow datasets together with the catchment selection and statistical analyses adopted. Section 5.3 reports and discusses key results regarding GHM performance and its implication for the projection of changes over time in the magnitude of extreme flows. Section 5.4 highlights the main findings and concludes with remarks on the scope to further improve understanding of changes in flood hazard at the global scale.

5.2 Data and methods

5.2.1 Observed and simulated streamflow datasets

The Global Streamflow Indices and Metadata (GSIM) archive (Do et al., 2018b, Gudmundsson et al., 2018b) is used as daily observational discharge for this analysis. Daily streamflow simulations available through the Inter-Sectoral Impact Model Intercomparison Project phase 2a and 2b (historical simulations spanning from 1971 to 2005; future simulations covering 2006-2099 period) are used, which are freely accessible at http://www.isimip.org. Six GHMs are considered: H08 (Hanasaki et al., 2008a, Hanasaki et al., 2008b), LPJmL (Von Bloh et al., 2010, Rost et al., 2008), MPI-HM (Stacke and Hagemann, 2012), ORCHIDEE (Guimberteau et al., 2014), PCR-GLOBWB (Van Beek et al., 2011, Wada et al., 2011), and WaterGAP2 (Müller Schmied et al., 2014, Mueller Schmied et al., 2016). To assess the model structure across GHMs, trends in streamflow extremes simulated under observational atmospheric forcing, available through the Global Soil Wetness Project Phase 3 – herein referred to as GSWP3 (Kim, 2017), were compared to observed trends. The influence of the acknowledged high uncertainty in climate model (Kumar et al., 2013, Kiktev et al., 2003) to streamflow simulations was assessed by comparing observed trends and trends simulated under 'hindcast' modelled atmospheric forcings, which were generated from four General Circulation Models (GCMs). To quantify the implication of model uncertainty to developing projections of flood hazard, trends simulated under projected modelled climate (using four GCMs) were also assessed. As a result, four simulation settings were used in this study, denoted by the atmospheric forcing and an overview is given in Table 5-1. These settings comprise two historical runs (GSWP3 and GCMHIND), and two future runs (GCMRCP26 and GCMRCP60), collectively leading to a total of 69 simulations.

For GSWP3 simulations, naturalised runs (i.e. human water management not taken into account) were chosen, since it enables the use of more GHMs when compared to the human impact runs (i.e. human water management inputs were used). A preliminary analysis also showed that trends in streamflow extremes exhibited from 'naturalised runs' are similar to those

obtained from 'human impact runs' (see section 4 of supplementary material) and thus this simulation choice is not likely to substantially affect the conclusions. Although significant efforts were made to keep the setting of each run as consistent as possible, there were nonetheless some differences owing to technical requirements across GHMs. As a result, the number of land grid points with available data is also slightly different across simulations.

Table 5-1: Summary of streamflow observation and simulations of this study

Reference window	Streamflow data	No. of sim	Description	Note
Historical (1971-2005)	GSIM	-	Observational streamflow selected from GSIM archive.	Streamflow observation for 3,666 unique locations
	GSWP3 (ISIMIP 2a)	6	Historical simulation forced by observational atmospheric forcing.	Model did not use human water management input.
	GCMHIND (ISIMIP 2b)	21	Historical simulation forced by 'hindcast' atmospheric forcing. Four GCMs were used: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and MIROC5.	No HadGEM2-ES simulation for MPI-HM. No HadGEM2-ES and MIROC5 simulations for ORCHIDEE.
Projection (2006-2099)	GCMRCP26 (ISIMIP 2b)	21	Future simulation forced by projected atmospheric forcing under socio-economic scenario RCP2.6. Four GCMs were used: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and MIROC5.	
	GCMRCP60 (ISIMIP 2b)	21	Future simulation forced by projected atmospheric forcing under socio-economic scenario RCP6.0. Four GCMs were used: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and MIROC5.	

5.2.2 Simulated streamflow extraction and catchment selection for observation-model comparison

To enable an observation-model comparison, simulated discharge needs to be extracted from gridded data products. Large-scale hydrological models, however, generally do not simulate discharge accurately over small-to-medium size catchments due to the coarse resolution of river network datasets (Hunger and Döll, 2008). To address this limitation, previous GHMs evaluations usually selected large catchments (a threshold of 9000km^2 was usually adopted, approximating the size of a one-degree longitude/latitude grid cell) and routed discharge (units: m^3/s) at the outlet of the catchment was used as simulated streamflow for a specific catchment (Zhao et al., 2017,

Veldkamp et al., 2018, Zaherpour et al., 2018, Liu et al., 2017). For evaluation studies that used relatively small catchments (e.g. area less than 9000km^2), the un-routed runoff simulation (units: mm/day) was extracted at the outlet while observed discharge was converted to runoff using catchment area prior to comparison (Gudmundsson et al., 2012b, Beck et al., 2017). To increase the sample size for the model-observation comparison (the first objective), the present study used both daily (i) un-routed runoff and (ii) routed discharge simulations, and thus two extraction procedures were adopted. A summary of these extraction procedure is provided below while detailed technical descriptions are provided in section 2 of supplementary material.

- For catchments with area from 0 to 9000km^2: un-routed runoff (mm/day) was extracted and then converted into discharge (m^3/s) by multiplying averaged runoff with catchment area. Specifically, catchment boundaries were superimposed on the GHM grid to obtain the weighted-area tables, which were then used to derive averaged runoff from the un-routed runoff simulation. For situations where there were two or more catchments having similar weighted-area tables (i.e. similar simulated streamflow would be extracted), runoff across these catchments were averaged (using catchment areas as weights) and the 'averaged time series' was used to represent all component catchments.
- For catchments with area greater than 9000 km2: the 'discharge output' approach (Zhao et al., 2017) was adopted to extract routed discharge (m3/s) from the GHM cell corresponding to each catchment.

To ensure sufficient data is available for historical trend analysis, only GSIM stations with at least 30 years available during the 1971-2005 period were considered (each year having at least 335 days of available records). As catchment boundary shapefiles were used to extract simulated streamflow for small catchments (Do et al., 2018a), stations were further filtered using two criteria: (i) availability of reported catchment area, and (ii) catchment boundary was accompanied with a "high" or "medium" quality flag (i.e. the discrepancy between reported and estimated catchment area is less than 10%).

A total of 4595 stations satisfied the quality selection criteria, of which large catchments (i.e. area greater than 9000km^2) with no suitable grid cell could be identified were further removed (11 catchments). For cases of two or more small catchments (i.e. area less than or equal to 9000km^2) having similar weighted-area tables, the 'averaged time series' (using catchment areas as weights) was calculated. A total number of 1542 time series fell in this category and were aggregated into 624 'averaged time series'. Figure 5-1 shows the spatial distribution of the final dataset for model-observation comparison, containing data for 3,666 locations (3,042 non-averaged time series and 624 averaged time series). The majority of available catchments are located in North America and Europe, with some data coverage improvements (relative to the streamflow database hosted by the Global Runoff Data Centre) over Asia and South America.

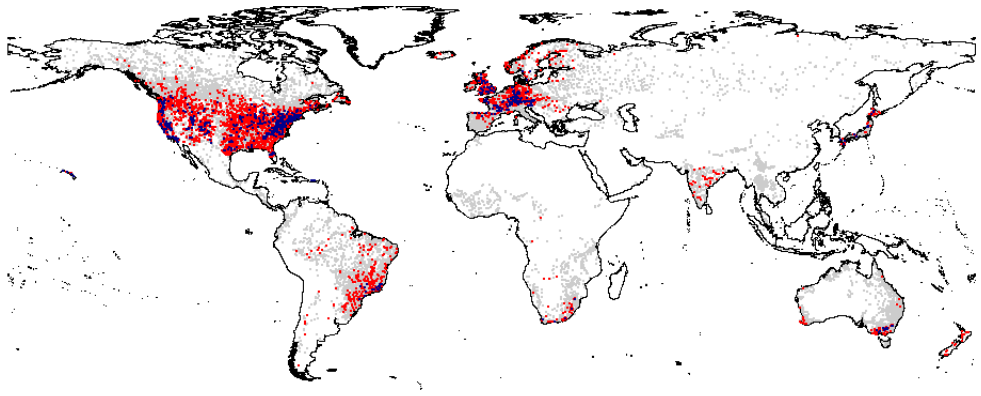


Figure 5-1: Locations of 3,666 streamflow observations (red dots: 3,024 non-averaged time series; blue dots: 624 averaged time series–where geographical coordinates were averaged from all component gauging coordinates) selected from GSIM archive for the model-observation comparison. Grey dots indicate GSIM time series that were removed due to insufficient data availability or quality.

5.2.3 Detecting trends in annual maximum streamflow

For each streamflow dataset, daily discharge was smoothed to 7-day averages to reduce variability in simulated streamflow, which can arise from the coarse routing parameters of GHMs (Dankers et al., 2014). The annual maximum time series of 7-day averaged discharge (labelled as the MAX7 index in the GSIM archive) was then derived to represent peak flow events. For gridded datasets, the centre averaged approach (e.g. averaged streamflow of Jan 7th is the mean value of Jan 4 – 10th) was used (default setting of the CDO software, freely available at https://code.mpimet.mpg.de/projects/cdo),

and the MAX7 timeseries was therefore derived for each GSIM station using this same approach. As a result, the derived value of the MAX7 index is slightly different from the value available in the online version of GSIM (Gudmundsson et al., 2018a), which applied the 'backward-moving average' technique (e.g. averaged streamflow of Jan 7th is the mean value of Jan 1 – 7th).

The magnitude of trends in the MAX7 index at a specific catchment or grid cell was quantified using the normalised Theil-Sen slope (Gudmundsson et al., 2019, Stahl et al., 2010) and the results are expressed in % change per decade. The significance of the at-site trend was assessed using a Mann-Kendall test at the 10% two-sided significance level (Wilks, 2011). The null hypothesis (no trend) is rejected if the two-sided p-value of the test statistic (Kendall's τ) is lower than 0.1, while the direction of the trend (i.e. increasing or decreasing) was determined using the sign of τ.

5.2.4 Statistical techniques to address three research objectives

To address three identified objectives, trends in streamflow extremes obtained from GSIM (observed trends) and ISIMIP simulations (simulated trends) are analysed. The at-site observed trends were available for 3666 observation locations while simulated trends were available for (i) 3666 observation locations (estimated from extracted discharge – described in section 5.2.2) and (ii) all GHM grid cells (estimated from routed discharge of each grid cell).

5.2.4.1 A hypothesis-test approach for comparison of trend characteristics

A range of hypothesis tests (summarised in Table 5-2; GSWP3 simulations were used to assess GHM uncertainty while GCMHIND simulations were used to assess the combined GCM-GHM uncertainty) was applied to address the first two objectives, which require comparing trend characteristics exhibited from different streamflow datasets. Four characteristics of trends were assessed:

- Trend mean: The mean (% change per decade) of trends in streamflow extremes across all gauge-/cell-based time series over a spatial domain. The hypothesis test was adopted to assess whether the trend

means exhibited from two specific streamflow datasets are significantly different.

- Trend standard deviation: The standard deviation (% change per decade) of trends in streamflow extremes across all gauge-/cell-based time series over a spatial domain. The hypothesis test was adopted to assess whether the trend standard deviations exhibited from two specific streamflow datasets are significantly different.
- Percentage of significant trends (%): The percentage of trends in a domain that are statistically significant, with gauge- or cell-based significance calculated using the Mann-Kendall test at the 10% significance level. The hypothesis test (field significant test) was adopted to assess whether the percentage of significant (increasing/decreasing) trends exhibited from a specific streamflow dataset is produced by random chance.
- Trend spatial structure: The spatial distribution of trends in streamflow extremes over a spatial domain. Pearson's correlation was used as a measure of similarity in trend spatial structure. The hypothesis test (pattern similarity test) was adopted to assess whether (i) the correlation between simulated trends introduced by GHMs and observed trends is significantly higher than zero, and (ii) the correlation between trends simulated under hindcast atmospheric forcing and observed trends is significantly lower than that between trends simulated under observational atmospheric forcing and observed trends.

Table 5-2: Hypothesis tests conducted to address the first two objectives

Objective	Null-Hypotheses	Streamflow dataset	Statistical tests
Objective 1: Capacity of GHMs to reproduce observed trends in flood hazards	Hypothesis 1: Trend means obtained from two streamflow datasets over observation locations were not statistically different.	(i) Observed discharge across 3666 observation locations (ii) Simulated discharge across 3666 observation locations (extraction processes outlined in Section 5.2.2)	Two-sample *t*-test at the 10% two-sided significance level
	Hypothesis 2: Trend standard deviations obtained from two streamflow datasets over observation locations were not statistically different.		Two-variance *F*-test at the 10% two-sided significance level
	Hypothesis 3: Percentage of significant trends obtained from all observation locations of a specific streamflow dataset was not produced by random chance.		Field significance test similar to that presented in Do et al. (2017) was adopted. A moving-block-bootstrap (block-length $L = 2$) was used to derive a null-hypothesis distribution of the change that occurred due to random chance. The null hypothesis is rejected when the true percentage falls on the right-hand side of the 95th percentile of the resampled distributions.
	Hypothesis 4: The correlation between trends obtained from two streamflow datasets was not significantly higher than '0' (i.e. zero pattern similarity).		'Zero pattern similarity' was compared to the probability distribution function (PDF) of pairwise correlation between simulated and observed trends, drawn from a bootstrap procedure similar to that proposed by Kiktev et al. (2003). The null hypothesis is rejected at 5% one-sided level when zero correlation falls on the left-hand side of the 5th percentile of the resampled distributions.
	Hypothesis 5: The correlation between GCMHIND simulated trends and observed trends was not significantly lower than the correlation between GSWP3 simulated trends and observed trends		The actual pairwise correlation between GCMHIND simulated trends and observed trends (denoted by $r_{GCMHIND}$) was compared to the bootstrapped PDF of correlation exhibited from GSWP3 simulated trends (denoted by r^*_{GSWP3}). If $r_{GCMHIND}$ falls on the left-hand side of the 5th percentile r^*_{GSWP3}, there is evidence to reject the null-hypothesis at the 5% one-sided significance level.
Objective 2: The representativeness of observation locations in the GHM simulations	Hypothesis 6: Trend mean obtained from observation locations was not statistically different from that obtained from all grid cells.	(i) Simulated discharge across 3666 observation locations (extraction processes outlined in Section 5.2.2) (ii) Routed discharge across all landmass grid cells (61,708 cells)	Two-sample *t*-test at the 10% two-sided significance level
	Hypothesis 7: Trend standard deviation obtained from observation locations was not statistically different from that obtained from all grid cells.		Two-variance *F*-test at the 10% two-sided significance level
	Hypothesis 8: Percentage of significant trends obtained from all grid cells of a specific streamflow dataset was not produced by random chance.		Field significance test similar to that presented in Hypothesis 3 but trends obtained from all grid cells were the subject of the assessment.

5.2.4.2 Estimating uncertainty of trend characteristics across ensemble members

The third and final objective, which focused on the implications of GCM-GHM uncertainty on projected changes in flood hazard, was addressed by quantifying the spread of trend characteristics (i.e. trend mean, trend standard deviation, and percentage of significant trends) exhibited from routed discharge projections under two representative concentration pathways.

The spatial uncertainty of projected trends (GCMRCP26 and GCMRCP60) was also quantified by calculating intra-/inter-model correlation of the trend patterns across all ensemble members available under the two projectios. Intra-model correlation represents spatial uncertainty introduced by GCM and was calculated across all pairs of simulated trends introduced by the same GHM (using different projected atmospheric forcing). Inter-model correlation represents the combined GCM-GHM spatial uncertainty, and was calculated for each pair of simulated trends (i) introduced by the different GHMs and (ii) forced with different projected climate data. This assessment also identified regions that were consistently detected with a significant increasing trend across most simulations, which can be used as an indication for potential 'hot-spots' of future flood hazard.

5.3 Results and discussion

5.3.1 Objective 1: Capacity of GHMs in reproducing observed trends in flood hazards

5.3.1.1 Model capacity at the global scale

Visual inspection of the normalised Theil-Sen slope across the GSIM time series shows a spatial pattern consistent with previous findings on trends in observed flood magnitude (Mangini et al., 2018, Do et al., 2017, Mallakpour and Villarini, 2015, Gudmundsson et al., 2019, Burn and Whitfield, 2018, Ishak et al., 2013). As shown in Figure 5-2 (top panel), decreasing trends tend to dominate Australia, the Mediterranean, western and north-eastern US and northern Brazil, while increasing trends appear mostly over central US, southern Brazil and northern Europe (including the UK). Across 96 stations with sufficient data quality in Asia (mostly located in Japan and southern

India), there are more stations showing decreasing trends than increasing trends.

The multi-model average of GSWP3 simulated trends (trends simulated under observational atmospheric forcing; middle panel of Figure 5-2) has generally good capacity to reproduce spatial patterns of trends in streamflow extremes at the continental scale. The multi-model average of GCMHIND simulated trends (trends simulated under hindcast atmospheric forcing; lower panel of Figure 5-2), however, could not reproduce important spatial associations of trends in streamflow extremes (e.g. the decreasing trends in south-eastern Australia, increasing trends over north-eastern Europe). The multi-model averages of both GSWP3 and GCMHIND simulations generally exhibit a lower magnitude of trends, reflecting the nature of averaging which tends to smooth out variability in trend magnitude across ensemble members. This feature is more prominent in GCMHIND (21 simulations available) compared to GSWP3 (six simulations available).

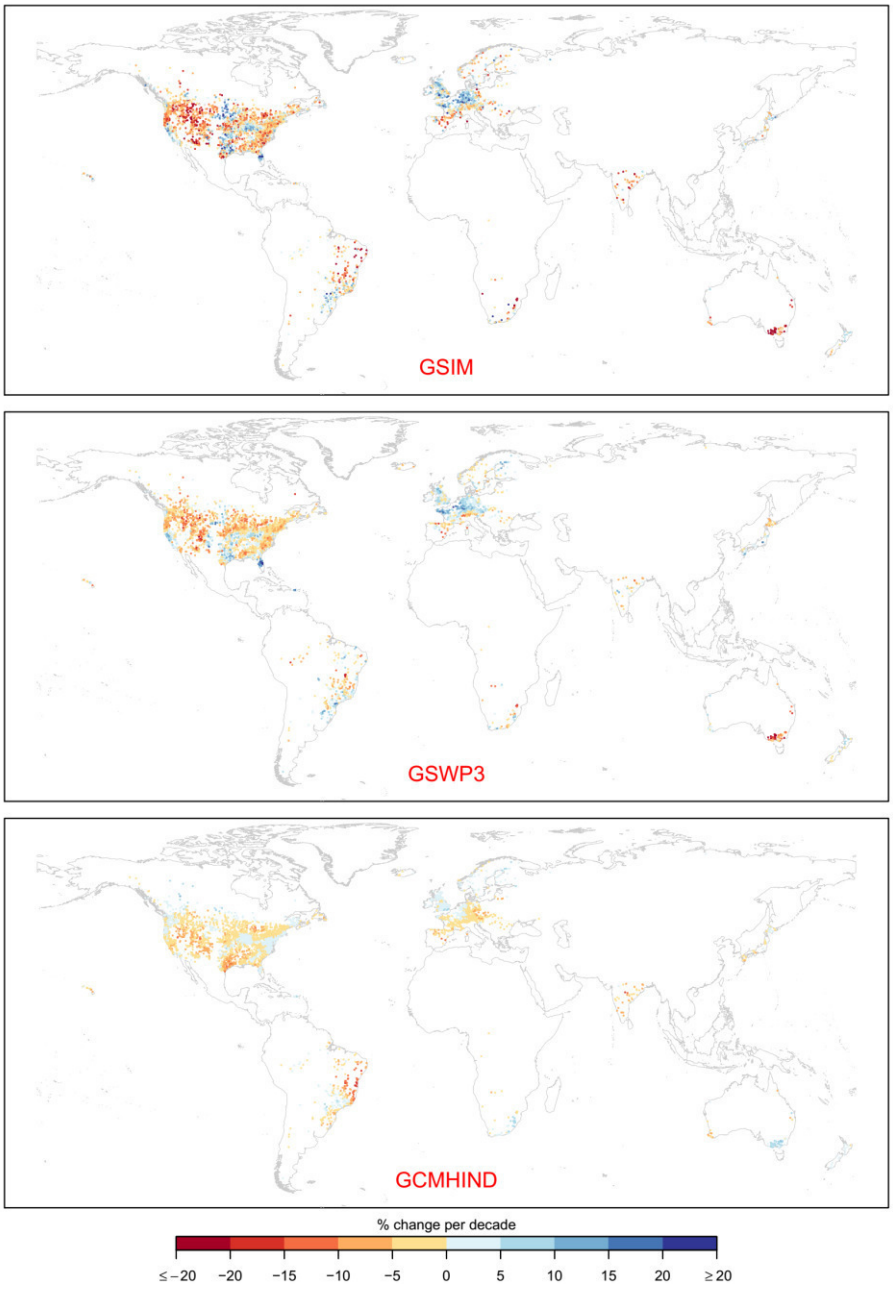


Figure 5-2: Normalised Theil-Sen slope for historical trends in flood magnitude (MAX7 index) exhibited over 3,666 locations across three streamflow datasets (top: GSIM; middle: GSWP3; bottom: GCMHIND). Multi-model average was showed for simulated trends. Trend is expressed in % change per decade.

To provide more detail on the influence of model uncertainty on simulated trends, four characteristics of trends in extreme flows (i.e. trend mean, trend standard deviation, percentage of significant trends and trend spatial structure) were assessed. The results are summarised in Table 5-3 for GSWP3 trends, and Table 5-4 for GCMHIND trends while GSIM results are showed in both tables for comparison. At the global scale, GSIM observed trends have the mean and standard deviation of -2.4% and 9.9% change per decade

respectively, with 7.5% (12.1%) locations showing significant increasing (decreasing) trends which is not field significant.

Table 5-3 shows the results of the model-observation comparison using GSWP3 simulated trends across six GHMs. Most simulated trends tend to have a significantly higher global mean (ranging from -2.2% to 0.1% change per decade) and lower standard deviation (ranging from 7.1% to 8.7% change per decade). The percentage of locations showing significant increasing/decreasing trends varies substantially across simulations, but field significance was not detected at the global scale. All GHMs possess moderate capacity in terms of simulating the spatial structure of trends (correlation ranges from 0.35 to 0.50), suggesting the appropriateness of current runoff generation concepts. There is, however, a notable difference in terms of the sign of trends simulated by different GHMs. Specifically, PCR-GLOBWB and WaterGAP2 tend to introduce a higher magnitude of trends compared to the other GHMs, evident through a higher global average (0.1% and -0.3% change per decade) and with a higher percentage of locations showing significant increasing trends (9.6% and 8.5% of locations) compared to percentage of locations showing significant increasing trends (6.1% and 4.2% of locations). This result suggests that using different GHMs can lead to different interpretations about changes in flood hazard at the global scale despite having a common boundary forcing. For example, simulated trends in streamflow extremes using PCR-GLOBWB suggest overall increasing trends at the global scale while trends introduced by LPJmL show more prominent decreasing trends.

Table 5-3: Characteristics of trends in the MAX7 index over the 1971-2005 period across 3666 locations for GSIM observed trends and GSWP3 simulated trends (six GHMs available). Trend mean and trend standard deviation are expressed in % change per decade. Correlation was obtained from GSIM observed trends and GSWP3 simulated trends for each GHM. Boldface texts represent values that reject the null-hypotheses outlined in Table 5-2.

GHM	Trend mean	Trend stand. dev.	% of sig. inc. trends	% of sig. dec. trends	Corr. obs. trend
GSIM (observation)	-2.4	9.9	7.5	12.1	-
H08	**-1.9**	**8.3**	4.8	6.7	**0.42**
LPJmL	-2.2	**7.1**	4.5	7.3	**0.37**
PCR-GLOBWB	**0.1**	**7.7**	9.6	6.1	**0.46**
WaterGAP2	**-0.3**	**8.2**	8.5	4.2	**0.49**
MPI-HM	-2.1	**8.7**	5.6	7.5	**0.50**
ORCHIDEE	**-1.4**	**8.6**	7	8.2	**0.35**

Table 5-4 provides the results of the model-observation comparison using GCMHIND simulated trends (intra-model averages are shown while results of individual simulations are reported in section 4 of supplementary material). Similar to GSWP3 trends, intra-model averages of GCMHIND trends tend to have a higher global mean (ranging from -2.3% to -0.4% change per decade with 19 out of 21 simulations suggesting a significantly different trend mean), lower trend standard deviation (ranging from 7.4% to 8.7% change per decade with all simulations suggesting a significantly different trend standard deviation). The composition between the percentages of locations showing significant trends varies substantially across simulations (ranging from 2.2%/4.4% to 12.2%/17.3% for significant increasing/decreasing trends) and field significance was found only for decreasing trends over three out of 21 simulations (two LPJmL simulations and one MPI-HM simulation). The multi-model ranges encapsulate the observed trend mean, and percentage of significant trends while the observed trend standard deviation is clearly above the range exhibited from all GCMHIND simulations.

Table 5-4: Characteristics of trends in the MAX7 index over the 1971-2005 period across 3666 locations for GCMHIND simulated trends. Trend mean and trend standard deviation are expressed in % change per decade. Intra-model averages of trend characteristics are shown for each GHM. Values in the parentheses show the number of simulations rejecting the null hypothesis outlined in Table 5-2 (out of four GCMs). Multi-model min/max/average values together with those exhibited from GSIM are also provided.

GHM	Trend mean	Trend stand. dev.	% of sig. inc. trends	% of sig. dec. trends	Corr. obs. trend
H08	-1.7 (4)	8.5 (4)	4.9 (0)	8.8 (0)	0.03 (2)
LPJmL	-2.3 (4)	7.9 (4)	4.2 (0)	12.6 (2)	0.09 (3)
PCR-GLOBWB	-1.1 (2)	7.4 (4)	7.5 (0)	9.4 (0)	0.06 (3)
WaterGAP2	-1.3 (4)	8.4 (4)	5.4 (0)	8.0 (0)	0.02 (2)
MPI-HM	-1.8 (3)	8.7 (4)	5.7 (0)	9.9 (1)	0.05 (2)
ORCHIDEE	-0.4 (2)	8.6 (2)	6.9 (0)	7.0 (0)	0.04 (1)
Multi-model min	-4.2	7.0	2.2	4.1	-0.06
Multi-model max	0.6	9.5	12.2	17.3	0.18
Multi-model average	-1.5	8.2	5.6	9.5	0.05
GSIM (observation)	-2.4	9.9	7.5	12.1	-

Among 21 GCMHIND simulations, ‘zero similarity’ hypothesis was rejected over 13 simulations, indicating that GCM-GHM ensemble members possess a capacity to simulate the spatial structure of observed trends in streamflow extremes. The correlation between GCMHIND simulated trends and GSIM observed trends (ranging from -0.06 to 0.18), however, is significantly lower than that exhibited from GSWP3 simulated trends across all GHMs (reported in Table 5-3). The results of similarity assessment are

illustrated for a single GHM (H08) in Figure 5-3, where the correlation between observed trends and GSWP3 simulated trends is significantly different from zero. In contrast, the correlation between observed trends and each of the simulated trends under hindcast atmospheric forcing (GCMHIND simulations) is much lower, with two of the four not being statistically higher than zero. These results confirm the substantial influence of atmospheric forcing uncertainty on the simulated trend pattern relative to GHMs structure.

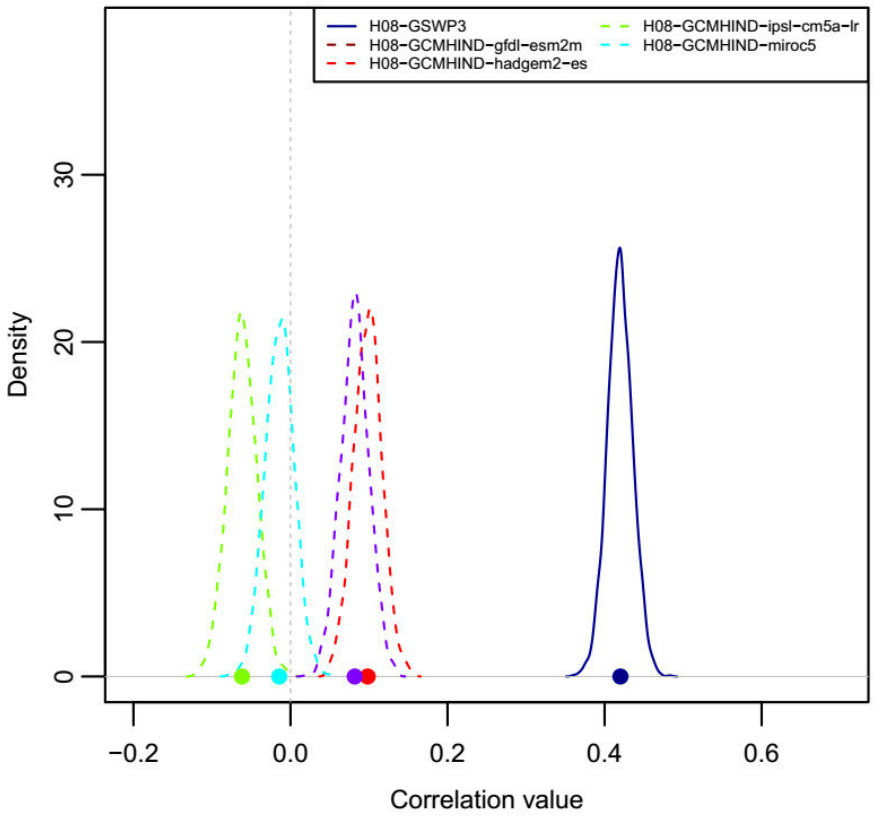


Figure 5-3: Model-observation correlation between observed trends and simulated trends across all simulations (GSWP3 and four GCMHIND simulations) of a single model (H08; similar results for other GHMs). Coloured dots indicate actual correlation between a specific simulated trend pattern and observed trend pattern across 3,666 locations. Colour lines represent the PDFs of correlation between simulated trend pattern and observed trend pattern obtained through a bootstrap resampling procedure (B = 2000).

5.3.1.2 Model capacity at the continental scale

Although saturation excess was used as the common framework to simulate runoff across all GHMs, different GHMs have different approaches to integrate components of the water balance (e.g. snow dynamics or soil moisture conditions), and thus model performance may vary across different regions. Assessing model capacity at the continental scale, therefore, can help diagnose factors of low performance of GHMs. For instance, performance of

climate models regarding trends in extreme rainfall (Kumar et al., 2013, Kiktev et al., 2007) may not be the key factor driving model performance over North America or Europe, as snowmelt also contributes substantially to flood generation (Cunderlik and Burn, 2004, Blöschl et al., 2017). On the other hand, South America is characterised with rainfall-induced floods and thus climate model uncertainty potentially plays a dominant role in model performance. The model-observation comparison (identical to that at the global scale) was therefore conducted over six continents and the results are summarised in Table 5-5 (multi-model averages are showed). Note that the observation locations are not evenly distributed (86% in North America and Europe), and thus the statistical power of continental assessment varies substantially across continents.

At the continental scale, the trend mean exhibited from GSIM ranges from -10.7% (Oceania) to 2.4% change per decade (Europe) while trend standard deviation ranges from 8.3% (Europe) to 15.8% change per decade (Oceania). The percentage of significant increasing (decreasing) trends exhibited from GSIM ranges from 3.2% to 22.6% (from 6.3% to 29.1%) and the composition of significant trends across the six continents is consistent with a previous investigation (Do et al., 2017). The observed percentage of significant trends is found to be above random chance for Europe (increasing flood magnitude) and Australia (decreasing flood magnitude) and this feature is captured quite well by GSWP3 simulated trends, with at least half of the simulations confirming field significances detected from GSIM.

Similar to the assessment at the global scale, most GSWP3 simulations generally exhibit a significantly higher trend mean (except for Africa) and lower trend standard deviation compared to observed trends, suggesting substantial uncertainty of trends in streamflow extremes introduced by GHMs at the continental scale. The spatial correlation is the weakest in Asia, as no simulation rejects the null-hypothesis of ‘zero similarity’, while the spatial correlation is the strongest in Oceania (correlation of 0.63). Oceania, however, exhibits the highest model-observation discrepancy in trend mean and trend standard deviation, indicating the capacity of a given GHM in terms

of the trend spatial structure is not necessarily consistent with its performance in terms of the mean and spread of trends.

GCMHIND simulations generally exhibit lower capacity in terms of reproducing trends, highlighting the presence of high uncertainty in GCM-GHM ensembles at the continental scale. Specifically, the majority of GCMHIND simulated trends tends to not capture the continental trend mean (except for Asia) and trend standard deviation (statistically different across all continents). GCMHIND trends also suggest the opposite composition between percentages of significant trends compared to GSIM trends over Europe, Africa and Oceania (e.g. simulated trends suggest more locations showing significant increasing trends while observed trends suggest the opposite) while the spatial correlation is significantly lower than GSWP3 correlation (except for Asia and South America). A notable feature is that the spatial correlation between GCMHIND trends and observed trends is the lowest in Oceania (correlation of -0.14), whereas GSWP3 suggested the strongest correlation in this continent, further indicating the substantial impact of atmospheric forcing relative to GHM structure on the ability to reproduce the spatial structure of trends in extreme flows.

Table 5-5: Characteristics of trends exhibited from GSIM/GSWP3/GCMHIND streamflow dataset at the continental scale (each observation location of 3,666 sites was allocated into one of the six continents). For simulated trends, only the multi-model average is shown for each region. Trend mean and trend standard deviation are expressed in % change per decade. Values in the parentheses show the number of simulations rejecting the null-hypothesis described in Table 5-2 (up to six for GSWP3 simulations and 21 for GCMHIND simulations). For GSIM, field significance of increasing/decreasing trends was highlighted by boldface texts.

	No. of	**Trend mean**			**Trend Stand. Dev.**			**% sig. inc. trends**			**% sig. dec. trends**			**Corr. obs. trends**	
Region	**loc.**	*GSIM*	*GSWP3*	*GCMHIND*	*GSIM*	*GSWP3*	*GCMHIND*	*GSIM*	*GSWP3*	*GCMHIND*	*GSIM*	*GSWP3*	*GCMHIND*	*GSWP3*	*GCMHIND*
Asia	96	-3.1	-1.2 (4)	-2.7 (6)	8.8	6.6 (5)	7.2 (15)	4.2	4.2 (0)	2.2 (0)	15.6	10.3 (1)	9.7 (2)	0.07 (0)	0.11 (11)
N. America	2441	-3.5	-2.4 (3)	-1.6 (18)	9.4	7.9 (6)	8.0 (19)	3.2	2.8 (0)	5.3 (0)	13.4	7.5 (0)	9.3 (3)	0.38 (6)	0.03 (12)
Europe	730	2.4	2.6 (6)	-0.7 (17)	8.3	7.1 (5)	5.9 (21)	**22.6**	20.2 (3)	7.3 (1)	6.3	2.1 (0)	10.1 (4)	0.43 (6)	0.10 (13)
Africa	48	-2.5	-1.3 (0)	1.5 (12)	14.8	9.8 (5)	8.0 (20)	6.3	2.8 (0)	9.6 (2)	10.4	10.4 (0)	3.3 (0)	0.46 (6)	0.07 (6)
S. America	265	-2.0	-0.2 (5)	-3.6 (14)	10.1	7.6 (6)	10.0 (20)	7.9	7.2 (0)	3.4 (1)	10.2	4.4 (0)	13.4 (5)	0.26 (6)	0.18 (17)
Oceania	86	-10.7	-6.1 (4)	2.4 (21)	15.8	10.9 (6)	8.4 (21)	4.7	3.7 (0)	11 (2)	**29.1**	22.1 (4)	1.9 (0)	0.63 (6)	-0.14 (2)

5.3.2 Objective 2: Determining the representativeness of observation locations in the GHM simulations

Observational evidence of changes in flood hazard is usually based on streamflow records over well-observed regions, and thus may not reflect the nature of change due to having a biased sample. Previous observation-based studies, as a result, usually addressed this issue by making inference only over 'data-covered regions' (Do et al., 2017, Ishak et al., 2013). To reconcile observed and simulated trends in historical streamflow extremes at the global scale, it is necessary to have trends of observation locations being a representative sample of those obtained from all land areas. This requirement is particularly important for sparsely observed continents (e.g. Asia, Australia, and Africa) to ensure comparable trends are used.

To assess the representativeness of observations locations in GHM grid cells, trend characteristics obtained from all simulated grid cells were compared to those estimated from the observation locations (3666 sites globally). The results, summarised in Table 5-6 (multi-model averages shown), suggest a significant difference between trend characteristics exhibited from all grid cells compared to those obtained from the observation locations for GSWP3 simulations. This feature is consistent at both global and continental scales, including North America and Europe – the continents with the best stream-gauge density. Specifically, the trend mean tends to get closer to zero, while the trend standard deviation obtained from all grid cells tends to be higher than that over observation locations. The difference between the percentages of significant increasing/decreasing trends across all grid cells also gets smaller. For instance, the percentage of locations with observations showing significant increasing (decreasing) trends over Oceania is 3.7% (22.1%) for GSWP3 multi-model averages (reported in Table 5-5), while the corresponding values are 10.7% (15.1%) over all grid cells (reported in Table 5-6). Finally, field significance for increasing/decreasing trends is detected in two/four out of six simulations, while the same feature could not be detected over the observation locations. These findings indicate that trends exhibited from observation locations are not a representative sample of trends obtained from all simulation grid cells and thus a common

model-observation picture of changes in global flood hazard remains elusive. To enable a holistic perspective of changes in extreme flows, it is therefore crucial to improve data accessibility and expand streamflow observational networks to ensure unbiased samples are available for large scale investigations.

The findings using GCMHIND simulations are similar in terms of the trend mean (closer to zero) and trend standard deviation (higher) across all grid cells relative to the observation locations. Across all land areas, the composition of the percentages of land mass showing significant trends exhibited by GCMHIND simulations contradicts that obtained from GSWP3 simulations for many continents. For example, GSWP3 simulations suggest more land areas showing significant decreasing trends than increasing trends over Asia and Oceania while GCMHIND simulations indicate an overall increasing change in extreme flows over the same continents. This feature further confirms the high uncertainty in the spatial structure of trends simulated under modelled climate data, which will be explored further in the next section.

Table 5-6: Characteristics of simulated trends across all grid cells at both continental and global scales (multi-model averages are showed). For each simulation, cell-based trend mean/trend standard deviation was compared to that of gauge-based trends (reported in Table 5-4). Values in parentheses represent the number of simulations reject the null-hypothesis described in Table 2 (up to six simulations for GSWP3 and 21 simulations for GCMHIND).

	Trend mean		**Trend Stand. Dev.**		**% sig. inc. trends**		**% sig. dec. trends**	
Region	*GSWP3*	*GCMHIND*	*GSWP3*	*GCMHIND*	*GSWP3*	*GCMHIND*	*GSWP3*	*GCMHIND*
Asia	-0.7 (3)	0.4 (16)	10.3 (6)	9.0 (15)	7.7 (0)	9.6 (7)	9.4 (3)	7.7 (4)
N. America	-0.9 (5)	0.9 (21)	10.4 (6)	8.6 (15)	8.1 (1)	9.6 (7)	10.7 (3)	5.9 (0)
Europe	1.1 (5)	0.2 (16)	8.5 (5)	8.4 (20)	11.5 (2)	9.1 (5)	4.5 (0)	7.9 (3)
Africa	0.7 (2)	-1.7 (15)	11.0 (3)	10.1 (12)	10.9 (1)	8.5 (6)	11.2 (2)	15.5 (11)
S. America	-2.0 (6)	-0.7 (19)	8.7 (3)	9.1 (17)	4.9 (0)	5.0 (0)	8.6 (0)	8.2 (1)
Oceania	-1.0 (6)	0.5 (17)	11.4 (4)	10.4 (17)	10.7 (0)	10.3 (3)	15.1 (1)	9.6 (6)
Global	-0.5 (6)	0.0 (20)	10.3 (6)	9.4 (19)	8.5 (2)	8.9 (8)	8.9 (4)	8.7 (4)

5.3.3 Objective 3: The implication of simulation uncertainty on the projection of trends in flood hazard

This section focuses on the uncertainty in simulated trends under projected atmospheric forcing at the global scale and its implication on identifying high-risk regions (i.e. robustly projected with significant increasing trends) in the context of climate change. Projections of routed discharge were used to derive trends in the MAX7 index across all grid cells under two socio-economic scenarios (i.e. RCP26 and RCP60). Trend mean,

trend standard deviation and percentages of significant trends were estimated for each ensemble member, and the spreads of these metrics were then assessed. In addition, the spatial uncertainty was also assessed, using correlations across all pairs of trend patterns projected by the same GHM (intra-model) and different GCM-GHM ensemble members (inter-model). For MPI-HM, streamflow was only simulated across the stream-network (approximately 45% of the global land grid cells), and thus three simulations of this GHM were removed from the analysis. As a result, only 18 ensemble members were used to explore the uncertainty in projected trends (GCMRCP26 and GCMRCP60 – trends estimated for the 2006-2099 period).

Table 5-7 shows a relatively low spread of the global trend mean (ranging from -1.1% to 1.0% change per decade; multi-model average of 0.1% and 0.2% change per decade for GCMRCP26 and GCMRCP60 respectively) and trend standard deviation (ranging from 1.8% to 4.4% change per decade) across ensemble members. LPJmL and ORCHIDEE generally suggest a decreasing trend at the global scale, evident through the negative global mean and more grid cells showing significant decreasing trends. The standard deviation of trends in future simulations (multi-model average of 2.3% and 3.4% change per decade for GCMRCP26 and GCMRCP60 respectively) is substantially lower than the historical run (multi-model average of 9.4% change per decade as reported in Table 5-6), potentially due to the capacity of longer time series in capturing the inter-decadal variability of the streamflow regimes, with both dry and wet periods being considered (Hall et al., 2014). The composition between percentages of grid cells showing significant trends varies substantially, ranging from 7.4% (7.1%) to 31.3% (33.6%) for significant increasing (decreasing) trends. Projected trends under the RCP26 scenario generally have closer to zero mean and lower standard deviation compared to those introduced by the RCP60 scenario, reflecting the nature of a ‘low-end warming’ scenario, when anthropogenic climate change reaches its peak at the middle of the 21st century followed by a generally stable condition.

Table 5-7: The uncertainty in the characteristics of projected trend (GCMRCP26 and GCMRCP60) across 18 members at the global scale (five GHMs). Trend mean and trend standard deviation have unit of %-change per decade. At-site significance of trend was identified using Mann-Kendall test at 10% level and the percentage of grid cells showing significant increasing/decreasing trends was reported. Intra-model average value of each metric across is shown for each GHM (numbers of simulations are provided).

Model	**No. of members**	**Trend mean**		**Trend standard deviation**		**% of significant increasing trends**		**% of significant decreasing trends**		**Intra-model correlation**		**Inter-model correlation**	
		GCM RCP26	***GCM RCP60***	***GCM RCP26***	***GCM RCP60***	***GCM RCP26***	***GCM RCP60***	***GCM RCP26***	***GCM RCP60***	***GCM RCP26***	***GCM RCP60***	***GCM RCP26***	***GCM RCP60***
H08	4	0.2	0.6	2.5	3.8	14.7	24.7	11.4	18.5	0.16	0.51	0.02	0.26
LPJmL	4	0.0	-0.1	2.1	3.1	10.3	20.7	9.2	19.0	0.05	0.43	0.02	0.22
ORCHIDEE	2	-0.4	-0.7	2.6	3.6	9.9	16.2	17.0	26.9	0.07	0.36	0.02	0.14
PCR-GLOBWB	4	0.1	0.1	2.4	3.4	15.9	24.5	11.4	19.4	0.07	0.32	0.03	0.20
WaterGAP2	4	0.3	0.6	2.3	3.2	13.5	27.6	7.8	11.4	0.04	0.32	0.02	0.21
Multi-model min	-	-0.5	-1.1	1.8	2.6	7.4	14.7	7.1	9.3	-0.02	0.14	-0.10	-0.17
Multi-model max	-	0.5	1.0	2.9	4.4	18.7	31.3	20.3	33.6	0.30	0.57	0.21	0.23
Multi-model average	-	0.1	0.2	2.3	3.4	13.2	23.5	10.7	18.2	0.08	0.39	0.01	0.06

Uncertainty in the spatial structure of trends in streamflow extremes is investigated using both intra-model (to reflect GCM uncertainty) and inter-model correlations (to reflect the combined GCM-GHM uncertainty). Over most GHMs, a more robust spatial pattern of projected trends under a 'business as usual' socio-economic scenario (RCP60) was found, indicated through higher intra-/inter-model correlation (multi-model averages of 0.39/0.06) compared to those exhibited from trends simulated under a 'low-end warming' socio-economic scenario (RCP26; multi-model averages of 0.08/0.01). The inter-model correlation (ranging from -0.17 to 0.23) is consistently lower than intra-model correlation (ranging from -0.02 to 0.57) due to the combined uncertainty of both GHMs and GCMs.

To quantity the robustness in terms of regions with significant trends in streamflow extremes, the number of simulations showing significant increasing/decreasing trends was counted at each grid cell (value ranging from 0 to 18). As showed in Figure 5-4, the projections under RCP26 (top panels) do not suggest many regions projected with an increasing trend by most ensemble members, but consistently suggest decreasing trends over the majority of Africa, Australia and the western America. Projections under the RCP60 scenario (lower panels) show substantially higher robustness in terms of regions with significant changes over time in streamflow extremes. Specifically, significant increasing trends are projected consistently over southern and south-eastern Asia, eastern Africa, Siberia and Greenland while high agreement of decreasing trends is found over southern Australia, north-eastern Europe, the Mediterranean and north-western North America. These findings share some similarity with a previous investigation that used ISIMIP fast track to identify regions projected with an increasing magnitude of 30-year return level of river flow (Dankers et al., 2014) as both studies suggest an overall (1) increasing trend over Siberia and South-East Asia, and (2) decreasing trend over north-eastern Europe and north-western North America. The present study, however, highlights a dominant decreasing trend over Australia, which was not shown previously. The different numbers of ensemble members (45 in Dankers et al. (2014) and 18 in the present study) and socio-economic scenarios (RCP8.5 in Dankers et al. (2014) and RCP2.6/RCP6.0 in the present study) between two studies indicate that the

choice of GCM-GHM ensemble and socio-economic scenarios could lead to substantially different projections of changes in flood hazard.

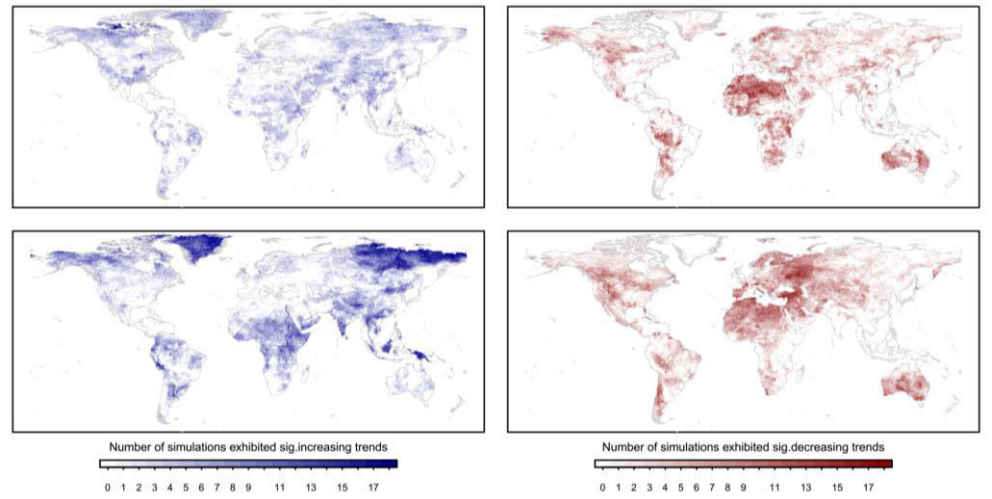


Figure 5-4: Number of simulations showing statistically significant trends at the 10% level at each grid cell. The left panels show results for the assessment of increasing trends, while the right panels show results for significant decreasing trends. Top: results of GCMRCP26 simulations; Bottom: results of GCMRCP60 simulations.

These results further confirmed the key role of GCM uncertainty in projecting changes in flood hazards, emphasising the importance of an agile adaptation strategy at the regional scale that can take into account the uncertain future of the Anthropocene (Dankers et al., 2014). Such a strategy is achievable only through a reliable and robust understanding of the change in flood hazard. The assessment of the representativeness of streamflow observations (section 5.3.2), however, demonstrated that the selected observation locations are not a representative sample of the entire land mass and thus inference of changes in flood hazard may be biased toward well-observed regions. To further quantify data coverage limitations, we used stream-gauge locations available in GSIM metadata, the to-date most comprehensive global databases of daily streamflow records, to identify the proportion of stations covering 'high-risk regions' (i.e. robustly projected with a significant increasing trends). Each grid-cell of the routed discharge simulation grid was categorised into one of five groups based on the number of GCMRCP60 simulations projecting a significant increasing trend (grouping criteria described in Figure 5-5). Each available stream-gauge location, of all the 30,959 GSIM stream-gauges, was then allocated into one of these five groups based on the gauge's geographical coordinates.

As showed in Figure 5-5, approximately 13.5% of the global land mass was projected with a significant increasing trend over at least 11 ensemble members (9.4% of Group 4, from 11-15 ensemble members projecting an increasing trend, and 4.1% of Group 5, from 16-18 ensemble members projecting an increasing trend). These 'likely increasing flood hazard' regions under a 'business as usual' scenario, however, were represented by less than 1% of all stations listed in the metadata of GSIM archive while approximately 90% of GSIM stations were located in low-risk regions (i.e. projected with a significant increasing trend in streamflow extremes by no more than five members). This indicates our limited understanding about the evolution of global flooding, which may lead to a risk of errors where increasing changes in flood hazard may have occurred in several ungauged regions.

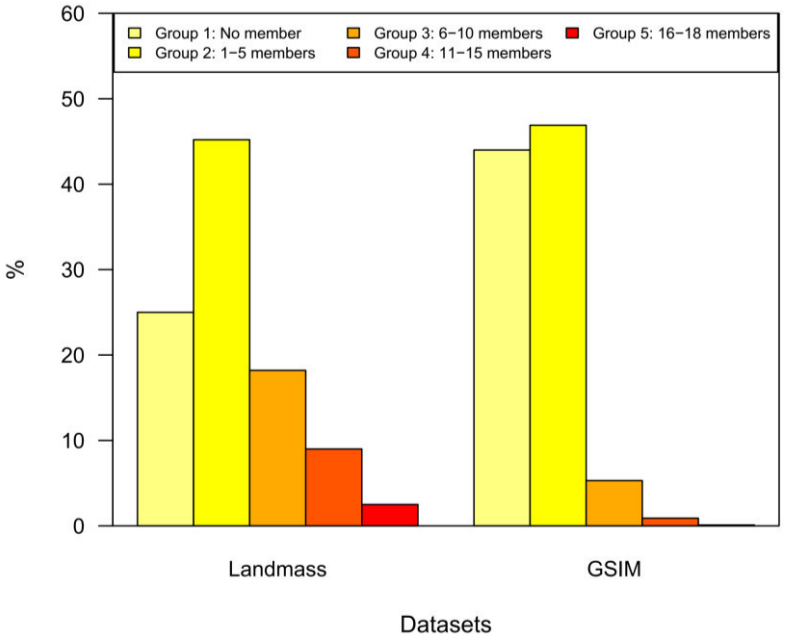


Figure 5-5: Percentage of each data-points grouped by the number of ensemble members projecting a significant increasing trend in seven-day annual maximum streamflow under RCP60 scenario. The range of possible numbers is from 0 to 18 and binned into five groups based on the number of members projecting a significant increasing trend (Group 1: no members, Group 2: from 1 to 5 members, Group 3: from 6 to 10 members, Group 4: from 11 to 15 members and Group 5: from 16 to 18 members).

5.4 Summary and conclusions

To reconcile observed and simulated trends in historical flood hazards at the global and continental scale, this study evaluated the capacity of six global hydrological models (GHMs) in terms of reproducing characteristics of

historical trends over the common 1971-2005 period, using the Global Streamflow Indices and Metadata (GSIM) archive as the observational basis. The observed trends in annual maximum streamflow confirm previous findings about changes in flood hazard over data-covered regions (Do et al., 2017), in which significant decreasing trends were found mostly in Australia, the Mediterranean, western US, eastern Brazil and Asia (Japan and southern India), while significant increasing trends were more common over central US, southern Brazil, northern Europe.

Several factors were assessed to identify the ability of GHMs to reproduce trends in streamflow extremes, focusing on four characteristics of trends (i.e. the mean and standard deviation of trends, the percentage of stations showing significant increasing/decreasing trends, and the spatial structure of trends). Simulated trends using an observational atmospheric forcing possess a moderate capacity to reproduce the characteristics of observed trends, confirming the appropriateness of GHM model structures for assessment of long-term changes in flood hazard. Atmospheric forcing uncertainty, however, significantly reduced the GHM's capacity in terms of simulating the spatial structure of trends, evident through significantly lower spatial correlation between observed trends and simulated trends using modelled atmospheric forcing. The simulated trends obtained from the subset of observation locations inadequately represented trends obtained from all GHM grid cells (at the continental and global scales). This was evident in most simulations for trend mean and trend standard deviation, indicating a potential mismatch between observation-based and model-based inferences about changes in flood hazard. To reconcile observed and simulated trends in streamflow extremes, it is critical to maintain and expand streamflow observation networks to provide representative samples for global scale investigations.

Uncertainties of trends in streamflow extremes were analysed to assess their implication on the development of projected changes in flood hazard over the common 2006-2099 period. Under both RCP26 and RCP60 socio-economic scenarios, simulated trends across ensemble members have relatively low uncertainty in terms of the global trend mean (ranging from -1.1% to 1.0% change per decade) and trend standard deviation (ranging from 1.8% to 4.4%

change per decade). The range of the percentage of land mass showing significant trends is high, ranging from 7.4% (7.1%) to 31.3% (33.6%) for significant increasing (decreasing) trends. The spatial correlations across inter-model trend patterns are generally low (ranging from -0.17 to 0.23), further indicating high levels of uncertainty. In terms of regional planning to mitigate flood hazard, individual models may provide contradictory signals of changes in flood hazard for a specific region. Under a 'business-as-usual' scenario (GCMRCP60), some regions, e.g. south-eastern Asia, eastern Africa, Siberia and Greenland, were consistently projected with significant increasing trends, which has some similarity to previous findings that used ISIMIP-Fast track simulations (Dankers et al., 2014). These 'high-risk' regions, however, are sparsely sampled, covered by less than 1% of all available stream-gauges listed in the catalogue of GSIM. Data coverage, as a result, remains the key limitation of this study, which could potentially lead to an erroneous conclusion in the state-of-understanding of trends in flood hazard, in which substantial changes, although having occurred, were not captured by available streamflow records.

Improved understanding of changes in flood hazard at global and continental scales, considering the many factors influencing model capacity, is achievable only through the combined efforts of many communities. Data providers, considering their tremendous investments in maintaining and making streamflow observations available in the public domain, remain key agencies to enhance the evidence-basis of the global terrestrial water cycle and changes in flood hazard. There are also limitations in GHMs that need to be addressed, but they can be complemented by progresses from other fields. The spread of trends in streamflow extremes (trend standard deviation) could be simulated more accurately by finer spatiotemporal resolution GHMs. Such an improvement in GHMs, however, is highly dependent on the improvement of input datasets (e.g. dam operations, historical irrigation databases and land-use/land-cover, in addition to atmospheric forcing), which are driven by advances in other geophysical disciplines (Bierkens et al., 2015, Wood et al., 2011). The moderate capacity of GHMs in terms of simulating the spatial structure of trends in streamflow extremes indicates the need for improved representation of runoff generation at the global scale (e.g. to better reflect rainfall-runoff relationship,

the complex effects of antecedent soil moisture, and the contribution of snow-dynamics). This research agenda is also a priority of large-sample hydrology (Gupta et al., 2014, Schaake et al., 2006), of which the focus is to develop generalisable models through investigating streamflow regimes over a large number of catchments. Large-sample GHM evaluations (Veldkamp et al., 2018, Zhao et al., 2017, Zaherpour et al., 2018) exploiting global databases such as the Global Runoff Data Base or GSIM (Gudmundsson et al., 2018a, Gudmundsson et al., 2018b, Do et al., 2018a, Do et al., 2018b) should also be emphasised to further quantify the reliability of GHMs according to different variables. Finally, uncertainty in GCMs, a long-standing challenge for the climate community, should also be addressed to enable robust projections of flood hazard in a warmer climate. One possibility is through constraining model performance using historical observations, which could potentially reduce the uncertainties of atmospheric forcing projections (Greve et al., 2018, Lorenz et al., 2018, He and Soden, 2016, Padrón et al., 2019).

Supplementary materials for Chapter 5

1 Simulation information

This section summarises the key simulation settings of each global hydrological model (GHM). Note that more detailed information is available in the protocols of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) available at https://www.isimip.org/protocol.

There are two important inputs for GHM simulations that were summarised in Table S5-1. Note that GSWPv3 was used as the sole observational atmospheric forcing dataset in this investigation beside modelled atmospheric forcing datasets introduced by four global climate models (GCM): GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and MIROC5.

1. Climate & CO_2 concentration scenarios (i.e. atmospheric forcing)
    - gswp3: observations-based dataset providing the climate forcing data.
    - rcp26: future climate and CO_2 concentration from RCP2.6.
    - rcp60: future climate and CO_2 concentration from RCP6.0.
    - hindcast: historical modelled climate and CO_2 concentration.
2. Human influence and land-use scenarios
    - nosoc: Naturalized runs (no human impact). No irrigation. No population and GDP data prescribed.
    - varsoc: Varying historical land use and other human influences.
    - 2005soc: Fixed year-2005 land use and other human influences.

Table S5-1. Simulations set up of GHMs used in this investigation. 'Climate' represents atmospheric forcing dataset while 'human' represents human influence and land-use scenarios

<table>
<tr><th>Model</th><th>GSWP3_VARSOC</th><th>GSWP3_NOSOC</th><th>GCMHIND</th><th>GCMRCP26</th><th>GCMRCP60</th></tr>
<tr><td>H08</td><td rowspan="4">Human: varsoc
Climate: gswp3</td><td rowspan="6">Human: nosoc
Climate: gswp3</td><td>Climate: hindcast
Human: 2005soc</td><td rowspan="5">Climate: rcp26
Human: 2005soc</td><td rowspan="5">Climate: rcp60
Human: 2005soc</td></tr>
<tr><td>LPJmL</td><td rowspan="4">Climate: hindcast
Human: varsoc</td></tr>
<tr><td>PCR-GLOBWB</td></tr>
<tr><td>WaterGAP2</td></tr>
<tr><td>MPI-HM</td><td rowspan="2">Simulations not available</td></tr>
<tr><td>ORCHIDEE</td></tr>
</table>

The results of preliminary assessment over 3666 observation locations suggest minor influence of human influence and land-use scenarios on the characteristics of trends in streamflow extremes (see section 4 of this supplementary material), and thus only GSWP3_NOSOC was used in the main text (denoted as GSWP3 in the main text).

2 Simulated streamflow extraction

2.1 Producing weighted-area tables for stations with catchment from 0 to 9000 km^2:

For stations with catchment area less than or equal to 9000 km2, catchment boundary was superimposed to ISIMIP grid to identify intersect cells and a weighted-area table was calculated for each case. Simulated runoff was extracted by averaging un-routed runoff from all intersect cells (considering weight). Runoff was then converted into discharge data.

Figure S5-1 provides an illustration of the weighted-area table for station US_0002282 (red dot) which has the total number of 15 upstream cells (dark-grey cells). Two components of the weighted-area table were used to label intersect cells: (1) cell number (dark red) and (2) fraction of each cell (weights) that is covered by the catchment boundary (dark blue). The weights were normalised such that they add up to one and used to extract simulated runoff for this catchment.

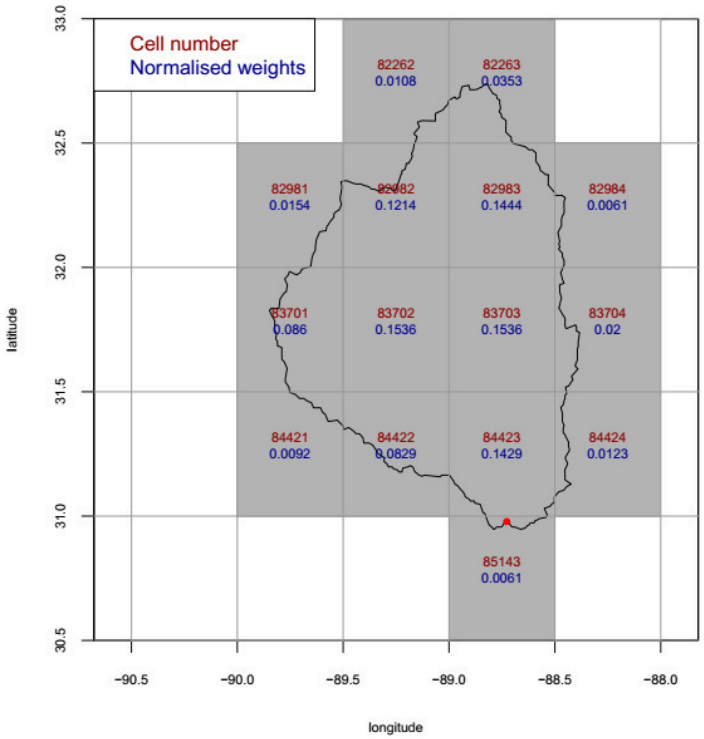
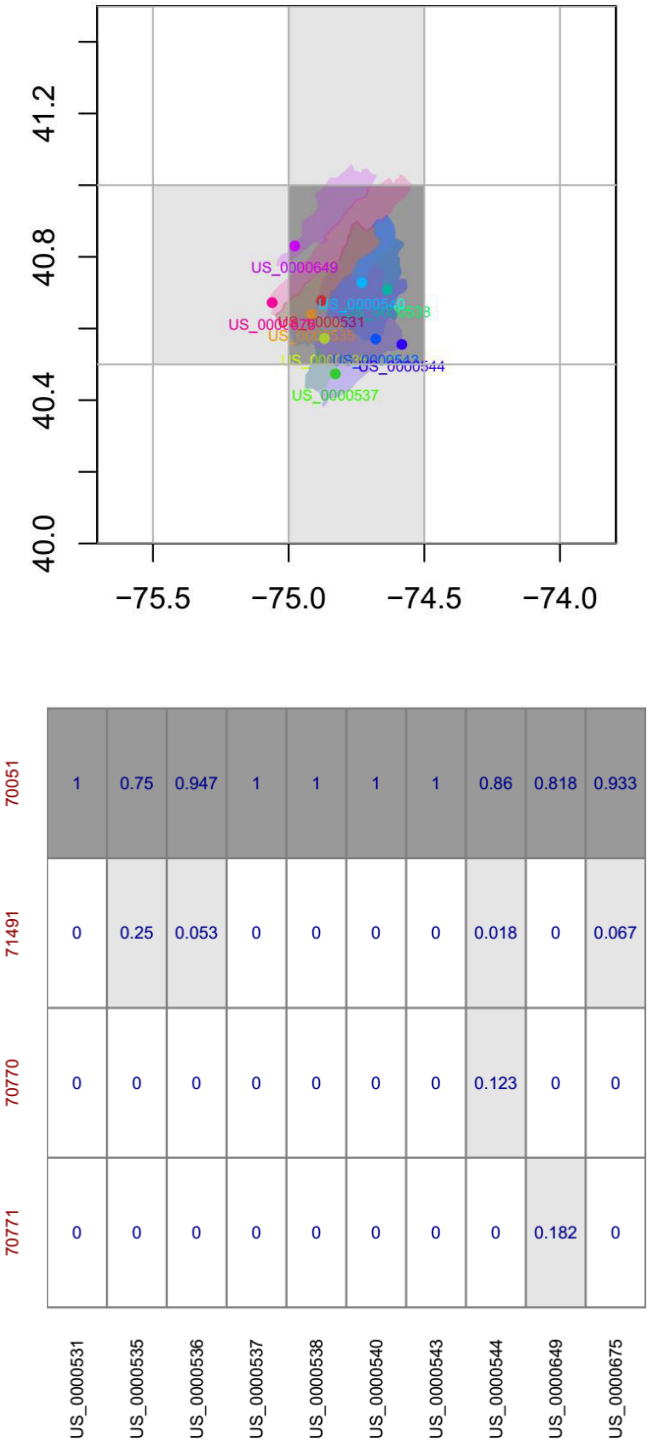


Figure S5-1. Illustration of the table of weights.

2.2 Averaging approach for cases where there were more than one catchment sharing similar weighted-area tables

Among catchments that have area less than 9000km2, there are many instances where two or more catchments have (almost) identical simulated runoff as they have similar weighted-area tables. All ISIMIP models have a common assumption of uniform runoff generation in the 0.5×0.5 grid area for the runoff generation, which in concept should represent an average value of runoff at finer resolution. Here we also treat catchments that intersect an identical set of dominant contributing grid-cells (total weights of at least 70%) as samples of an identical simulation domain. As a result, the area-weighted mean discharge of these catchments was calculated and used for model-observation comparison.

A search was conducted across all weighted-area tables to identify cases that have an identical set of intersecting cells contributing at least 70% to the total weighting. Figure S5-2 provides an example of these cases. In the top panel, boundaries of ten catchments were supperimposed on top of ISIMIP gridline (0.5×0.5 degree) and demonstrates that they share a common cell (number 70051) which contributes at least 70% to the total weight (showed in the bottom panel).



	US_0000531	US_0000535	US_0000536	US_0000537	US_0000538	US_0000540	US_0000543	US_0000544	US_0000649	US_0000675
70051	1	0.75	0.947	1	1	1	1	0.86	0.818	0.933
71491	0	0.25	0.053	0	0	0	0	0.018	0	0.067
70770	0	0	0	0	0	0	0	0.123	0	0
70771	0	0	0	0	0	0	0	0	0.182	0

Figure S5-2. Example of instances where there are significant overlapping in contributing cells. Left: locations of 10 catchments that share a common contributing grid-cell (cell number 70051 (in dark-grey colour) contributes at least 70% to the total weight of each catchment) although specific catchment have different contributing cells. Right: detail information of weighted-area table of these 10 catchments.

Figure S5-3 illustrates another case where three different catchments share two common cells (no. 76524 and 76525). These cells contribute 100%, 79.1%, and 76.4% to the weighted-area tables of catchment US_0001198, US_0001199,

and US_0001203 respectively. In both examples, the identified catchments were considered samples of the same modeling domain.

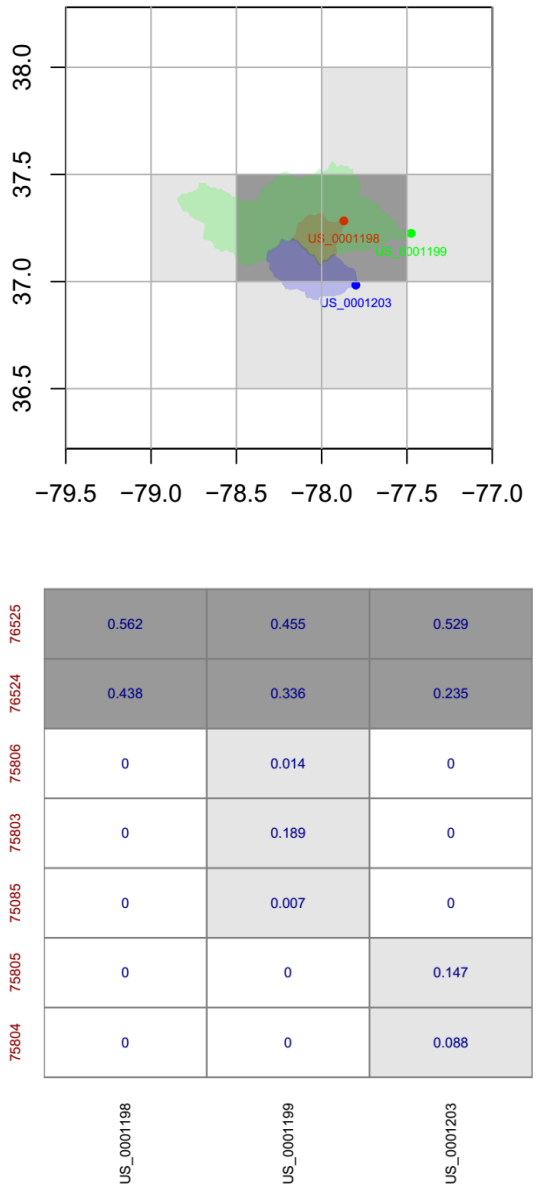


Figure S5-3. Similar to Figure S2, but here we have two contributing cells. The total weight of these common cells (number 76524 and 76525, highlighted in dark-grey colour) is higher than 0.7 in all cases and thus these three catchments were considered samples of the same modelling domain.

For these cases, average discharge $\bar{Q}$(m^3/s) was then calculated to represent these stations in model-observation comparison following below procedures:

For observation discharge:

1. Convert discharge Q (unit: m^3/s) to runoff rate R (unit: m/day) using catchment area A (unit: m^2) for each station.

$$R_i = Q_i \times 24 \times 3600 / A_i$$

(m/day)

Average catchment size was also recorded:

$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n} A_i \qquad (m^2)$$

2. Average runoff rate across all catchments (considering areal weights)

$$\bar{R} = \frac{\sum_{i=1}^{n} R_i A_i}{\sum_{i=1}^{n} A_i} \quad (m/day)$$

3. Back-calculate average discharge (m3/s):

$$\bar{Q} = \frac{\bar{R}\bar{A}}{24 \times 3600} \quad (m^3/s)$$

For simulation discharge:

1. Extract runoff rate using weighted-area tables as described in Section 2.1 for all catchments.
2. Follow Step 2 and Step 3 of the observation procedure.

2.3 Discharge output identification for catchment with area greater than 9000 km²

For catchments with area greater than 9000km^2, the 'discharge output' approach was adopted to find GHM cells corresponding to the catchments following Zhao et al. (2017). This process identified grid cell with the closest match between catchment area available in DDM30 routing network and the reported area for a specific catchment. The identified grid cell was then used to extract simulated discharge available in ISIMIP data repository. Stations were removed if the procedure could not identify any DDM30 grid cell surrounding the reported geographical location with drainage area discrepancy less than 30% (see supplementary of Zhao et al. (2017) for detail).

3 Spatial uncertainty across simulated trends forced with different modelled atmospheric forcing

The assessment in section 5.3.3 of the main text suggests the combined GCM-GHM uncertainty have led to the presence of high uncertainty in terms of regions projected with significant trends in streamflow extremes, with an

identical region could be projected with an increasing trend by one member and a decreasing trend by another member. This feature is illustrated in Figure S5-4, which shows the spatial structure of projected trends in annual streamflow magnitude (trends in MAX7 index) of two ensemble members and the ensemble mean.

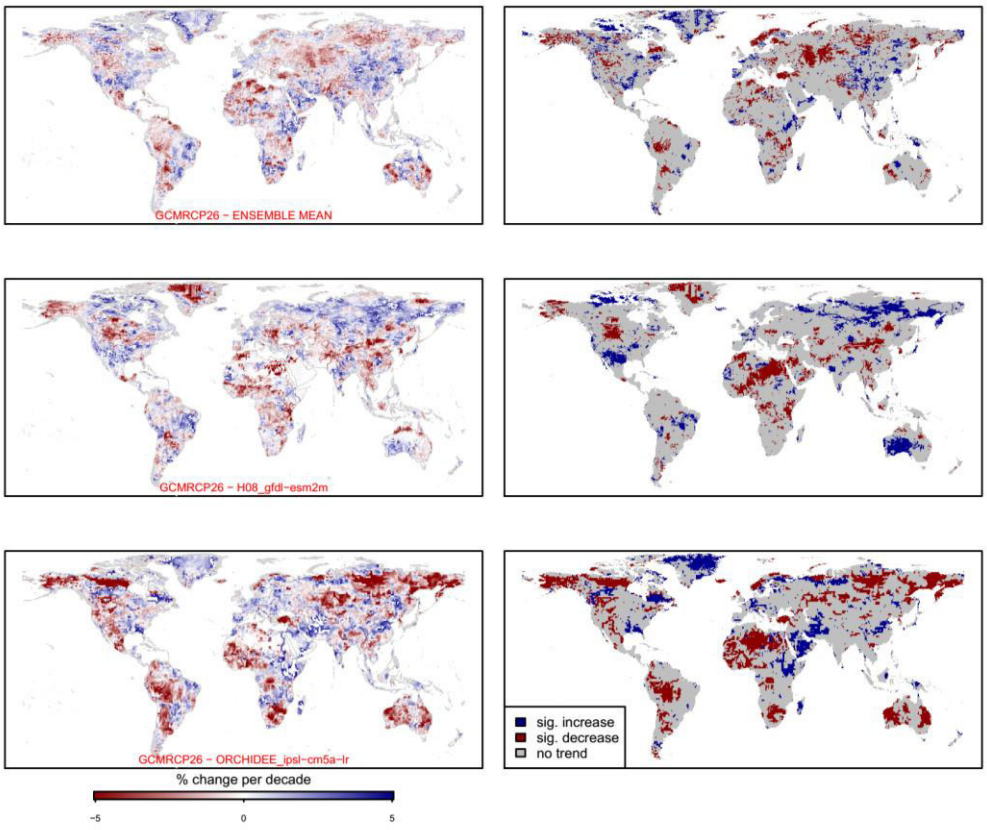


Figure S5-4. The magnitude (left panels) and significance (right panels) of trends in simulated MAX7 time series across all grid cells under RCP26 socio-economic scenario (2006-2099). The top panels show results of the GHMGCM ensemble mean (18 members); middle panels: H08 forced with gfdl-esm2m climate data; bottom panels: ORCHIDEE forced with ipsl-cm5a-lr climate data. These two models had the lowest value of pattern similarity.

Two simulations with the lowest spatial correlation under the RCP26 socio-economic scenarios (correlation of -0.17) provide a notable mismatch in terms of regions projected with significant increasing/decreasing trends in streamflow extremes. Specifically, H08 forced with GFDL-ESM2M (middle panels) projects an increasing trend for the majority of Australia and Siberia, while ORCHIDEE forced with IPSL-CM5A-LR projects an overall decreasing trend for the same regions. The ensemble mean (top panels) seems to take into account these differences and generates a combined picture of these two models. Note that the magnitude of trends (left panels of Figure S5-4) exhibited through the

ensemble mean is lower than the individual ensemble member due to the impact of averaging.

4 Supplementary Tables

Considering a large number of simulations available (73 in total), the main text mostly used multi-model min/max/average to illustrate the results for cases where there is more than one simulation available for an identical GHM/spatial-domain. For example, Table 5-3 of the main text (characteristics of trends in the MAX7 index over 1971-2005 period across 3666 locations globally) only showed the multi-model average for GCMHIND simulations of each GHM (up to four simulations available). This section, therefore, provides the results of each experiment at the global scale for individual models. Table S5-2 provides a list of all 73 available models reported in this section together with their simulation settings. Note that:

(i) GSWP3_VARSOC simulations (listed in Table S5-2 as H08_GSWVAR, LPJ_GSWVAR, PCR_GSWVAR, and WAT_GSWVAR) were not reported in the main text as (1) there were only four simulations available (comparing to six simulations of GSWP3_NOSOC) and (2) the results obtained from GSWP3_NOSOC and GSWP3_VARSOC are similar (Table S5-3).

(ii) In the main text, GSWP3_NOSOC simulations were denoted as GSWP3.

Table S5-2. Available ISIMIP streamflow simulations and associated setting.

Seq	Streamflow simulations	GHM	Climate	Human	Period
1.	H08_GSWVAR	H08	Observation (GSWPv3)	varsoc	1971-2005
2.	H08_GSWNO		Observation (GSWPv3)	nosoc	
3.	H08_HIN_G		hindcast (GFDL-ESM2M)	2005soc	
4.	H08_HIN_H		hindcast (HadGEM2-ES)		
5.	H08_HIN_I		hindcast (IPSL-CM5A-LR)		
6.	H08_HIN_M		hindcast (MIROC5)		
7.	H08_RCP26_G		rcp26 (GFDL-ESM2M)		2006-2099
8.	H08_RCP26_H		rcp26 (HadGEM2-ES)		
9.	H08_RCP26_I		rcp26 (IPSL-CM5A-LR)		
10.	H08_RCP26_M		rcp26 (MIROC5)		
11.	H08_RCP60_G		rcp60 (GFDL-ESM2M)		

Seq	Streamflow simulations	GHM	Climate	Human	Period
12.	H08_RCP60_H		rcp60 (HadGEM2-ES)		
13.	H08_RCP60_I		rcp60 (IPSL-CM5A-LR)		
14.	H08_RCP60_M		rcp60 (MIROC5)		
15.	LPJ_GSWVAR	LPJmL	Observation (GSWPv3)	varsoc	1971-2005
16.	LPJ_GSWNO		Observation (GSWPv3)	nosoc	
17.	LPJ_HIN_G		hindcast (GFDL-ESM2M)	varsoc	
18.	LPJ_HIN_H		hindcast (HadGEM2-ES)		
19.	LPJ_HIN_I		hindcast (IPSL-CM5A-LR)		
20.	LPJ_HIN_M		hindcast (MIROC5)		
21.	LPJ_RCP26_G		rcp26 (GFDL-ESM2M)	2005soc	2006-2099
22.	LPJ_RCP26_H		rcp26 (HadGEM2-ES)		
23.	LPJ_RCP26_I		rcp26 (IPSL-CM5A-LR)		
24.	LPJ_RCP26_M		rcp26 (MIROC5)		
25.	LPJ_RCP60_G		rcp60 (GFDL-ESM2M)		
26.	LPJ_RCP60_H		rcp60 (HadGEM2-ES)		
27.	LPJ _RCP60_I		rcp60 (IPSL-CM5A-LR)		
28.	LPJ_RCP60_M		rcp60 (MIROC5)		
29.	PCR_GSWVAR	PCR-GLOBWB	Observation (GSWPv3)	varsoc	1971-2005
30.	PCR_GSWNO		Observation (GSWPv3)	nosoc	
31.	PCR_HIN_G		hindcast (GFDL-ESM2M)	varsoc	
32.	PCR_HIN_H		hindcast (HadGEM2-ES)		
33.	PCR_HIN_I		hindcast (IPSL-CM5A-LR)		
34.	PCR_HIN_M		hindcast (MIROC5)		
35.	PCR_RCP26_G		rcp26 (GFDL-ESM2M)	2005soc	2006-2099
36.	PCR_RCP26_H		rcp26 (HadGEM2-ES)		
37.	PCR_RCP26_I		rcp26 (IPSL-CM5A-LR)		
38.	PCR_RCP26_M		rcp26 (MIROC5)		
39.	PCR_RCP60_G		rcp60 (GFDL-ESM2M)		
40.	PCR_RCP60_H		rcp60 (HadGEM2-ES)		
41.	PCR_RCP60_I		rcp60 (IPSL-CM5A-LR)		
42.	PCR_RCP60_M		rcp60 (MIROC5)		
43.	WAT_GSWVAR	WaterGAP2	Observation (GSWPv3)	varsoc	1971-2005
44.	WAT_GSWNO		Observation (GSWPv3)	nosoc	
45.	WAT_HIN_G		hindcast (GFDL-ESM2M)	varsoc	
46.	WAT_HIN_H		hindcast (HadGEM2-ES)		
47.	WAT_HIN_I		hindcast (IPSL-CM5A-LR)		
48.	WAT_HIN_M		hindcast (MIROC5)		
49.	WAT_RCP26_G		rcp26 (GFDL-ESM2M)	2005soc	2006-2099
50.	WAT_RCP26_H		rcp26 (HadGEM2-ES)		
51.	WAT_RCP26_I		rcp26 (IPSL-CM5A-LR)		
52.	WAT_RCP26_M		rcp26 (MIROC5)		
53.	WAT_RCP60_G		rcp60 (GFDL-ESM2M)		

Seq	Streamflow simulations	GHM	Climate	Human	Period
54.	WAT_RCP60_H		rcp60 (HadGEM2-ES)		
55.	WAT_RCP60_I		rcp60 (IPSL-CM5A-LR)		
56.	WAT_RCP60_M		rcp60 (MIROC5)		
57.	MPI_GSWNO	MPI-HM	Observation (GSWPv3)	nosoc	1971-2005
58.	MPI_HIN_G		hindcast (GFDL-ESM2M)	varsoc	
59.	MPI_HIN_I		hindcast (IPSL-CM5A-LR)		
60.	MPI_HIN_M		hindcast (MIROC5)		
61.	MPI_RCP26_G		rcp26 (GFDL-ESM2M)	2005soc	2006-2099
62.	MPI_RCP26_I		rcp26 (IPSL-CM5A-LR)		
63.	MPI_RCP26_M		rcp26 (MIROC5)		
64.	MPI_RCP60_G		rcp60 (GFDL-ESM2M)		
65.	MPI_RCP60_I		rcp60 (IPSL-CM5A-LR)		
66.	MPI_RCP60_M		rcp60 (MIROC5)		
67.	ORC_GSWNO	ORCHIDEE	Observation (GSWPv3)	nosoc	1971-2005
68.	ORC_HIN_G		hindcast (GFDL-ESM2M)	varsoc	
69.	ORC_HIN_I		hindcast (IPSL-CM5A-LR)		
70.	ORC_RCP26_G		rcp26 (GFDL-ESM2M)	2005soc	2006-2099
71.	ORC_RCP26_I		rcp26 (IPSL-CM5A-LR)		
72.	ORC_RCP60_G		rcp60 (GFDL-ESM2M)		
73.	ORC_RCP60_G		rcp60 (IPSL-CM5A-LR)		

Table S5-3. Characteristics of trends in the MAX7 index over the 1971-2005 period across 3666 locations introduced GHMs. Trend mean and trend standard deviation have unit of %-change per decade. Gauge-based significant trends were identified using a Mann-Kendall test (10% two-sided significance level). The global significance of this result is then calculated using field significance test (highlighted in boldface text). Trend mean, trend standard deviation and trend spatial structure were compared against that obtained from trends exhibited by GSIM (significant values were represented in boldface text).

Streamflow simulations	Trend mean	Trend standard deviation	Percentages of significant		Correlation against observed trends
			Increasing trend	Decreasing trend	
H08_GSWVAR	-2.0	**8.3**	4.8	6.7	**0.4**
LPJ_GSWVAR	-2.6	**7.5**	4.6	9.2	**0.4**
PCR_GSWVAR	**0.0**	**7.7**	9.4	6.1	**0.5**
WAT_GSWVAR	**-0.7**	**8.5**	8.4	5.8	**0.5**
H08_GSWNO	**-1.9**	**8.3**	4.8	6.7	**0.4**
LPJ_GSWNO	-2.2	**7.1**	4.5	7.3	**0.4**
PCR_GSWNO	**0.1**	**7.7**	9.6	6.1	**0.5**
WAT_GSWNO	**-0.3**	**8.2**	8.5	4.2	**0.5**
MPI_GSWNO	-2.1	**8.7**	5.6	7.5	**0.5**
ORC_GSWNO	**-1.4**	**8.6**	7	8.2	**0.4**
H08_HIN_G	**-0.4**	**8.9**	6.1	7.8	**0.1**
H08_HIN_H	**-2.8**	**8.4**	2.2	10.8	-0.1
H08_HIN_I	**0.1**	**8.9**	7.7	4.4	0.0
H08_HIN_M	**-3.6**	**7.8**	3.4	12.0	**0.1**
LPJ_HIN_G	**-0.8**	**8.0**	6.3	8.3	**0.1**
LPJ_HIN_H	**-2.9**	**8.1**	2.8	**14.6**	0.0
LPJ_HIN_I	**-1.3**	**8.0**	4.1	10.1	**0.1**
LPJ_HIN_M	**-4.1**	**7.3**	3.5	**17.3**	**0.2**
PCR_HIN_G	**-0.2**	**8.0**	8.3	9.0	**0.1**
PCR_HIN_H	-2.5	**7.1**	2.7	11.0	0.0
PCR_HIN_I	**0.6**	**7.6**	12.2	4.1	0.0
PCR_HIN_M	-2.1	**7.0**	6.9	13.5	**0.1**
WAT_HIN_G	**0.2**	**9.2**	8.2	5.6	**0.1**
WAT_HIN_H	**-2.9**	**8.1**	2.7	10.9	-0.1
WAT_HIN_I	**0.5**	**8.8**	6.2	4.2	-0.1
WAT_HIN_M	**-2.9**	**7.3**	4.3	11.4	**0.1**
MPI_HIN_G	**-1.3**	**9.5**	5.9	7.9	**0.1**
MPI_HIN_I	**0.2**	**9.2**	8.8	5.6	0.0
MPI_HIN_M	**-4.2**	**7.3**	2.3	**16.3**	**0.1**
ORC_HIN_G	**-0.9**	**8.6**	5.2	7.6	0.0
ORC_HIN_I	**0.1**	**8.6**	8.6	6.4	**0.1**

Table S5-4. Trend mean, trend standard deviation and percentage of significant trends exhibited from all simulation grid cells. Trend mean and trend standard deviation have unit of %-change per decade. Cell-based significance was identified using the Mann-Kendall test at the 10% significance level. The global significance of this result is then calculated using field significance test (highlighted in boldface text). Trend mean and trend standard deviation across all land mass were compared against that obtained across 3666 observation locations (reported in Table S5-3) and significant values are highlighted in boldface text.

Streamflow simulations	Trend mean	Trend standard deviation	Percentages of significant	
			Increasing trend	Decreasing trend
H08_GSWVAR	**-0.2**	**10.2**	8.7	10.3
LPJ_GSWVAR	**-1.4**	**10.3**	7.4	**13.5**
PCR_GSWVAR	**-1.0**	**10.9**	10.7	**14.7**
WAT_GSWVAR	**-0.1**	**11.3**	**11.4**	10.7
H08_GSWNO	**0.0**	**10.0**	8.6	9.2
LPJ_GSWNO	**-0.8**	**9.8**	7.6	**11.1**
PCR_GSWNO	**-0.9**	**10.8**	10.9	**14.4**
WAT_GSWNO	**0.3**	**11.1**	**11.5**	9.8
MPI_GSWNO	**-0.9**	**9.6**	5.9	**7.5**
ORC_GSWNO	**-0.4**	**10.4**	**6.7**	**7.4**
H08_HIN_G	**1.4**	**11.1**	15.6	10.4
H08_HIN_H	**0.2**	**8.7**	7.8	8.8
H08_HIN_I	**-0.3**	**9.8**	8.1	10.4
H08_HIN_M	**1.0**	**9.6**	10.8	7.7
LPJ_HIN_G	**-0.3**	**9.2**	8.6	8.8
LPJ_HIN_H	**-0.9**	**8.7**	5.2	9.5
LPJ_HIN_I	**-0.9**	**8.7**	6.2	8.9
LPJ_HIN_M	**-0.6**	**9.0**	7.9	9.1
PCR_HIN_G	**1.2**	**11.1**	14.7	10.8
PCR_HIN_H	**-0.3**	**8.6**	8.8	10.1
PCR_HIN_I	**-1.1**	**10.7**	8.7	11.7
PCR_HIN_M	**0.5**	**8.9**	12.1	9.7
WAT_HIN_G	**1.5**	**10.8**	14.8	**6.9**
WAT_HIN_H	**0.0**	**9.1**	6.1	7.0
WAT_HIN_I	0.3	**9.5**	7.7	7.2
WAT_HIN_M	**0.6**	**9.8**	11.2	6.9
MPI_HIN_G	**0.6**	9.5	8.1	**6.0**
MPI_HIN_I	**-0.9**	**8.2**	3.8	**6.5**
MPI_HIN_M	**-0.1**	7.4	4.3	**4.8**
ORC_HIN_G	**-0.2**	**10.2**	10.8	**10.8**
ORC_HIN_I	**-1.1**	**9.3**	6.2	10.7

Table S5-5. Characteristics of projected trend (GCMRCP26) across 18 members at the global scale. Mean and standard deviation have unit of %-change per decade. Cell-based significance was identified using the Mann-Kendall test at the 10% significance level. Note that no statistical test for comparison was conducted.

Streamflow simulations	Trend mean	Trend standard deviation	Percentages of significant	
			Increasing trend	Decreasing trend
H08_RCP26_G	0.0	2.1	10.5	10.2
H08_RCP26_H	0.5	2.7	18.7	10.5
H08_RCP26_I	0.1	2.4	12.5	13.6
H08_RCP26_M	0.1	2.9	17.2	11.2
LPJ_RCP26_G	-0.1	1.8	7.4	7.3
LPJ_RCP26_H	0.0	2.1	10.9	10.2
LPJ_RCP26_I	0.0	2.1	10.4	10.4
LPJ_RCP26_M	0.0	2.2	12.4	8.7
PCR_RCP26_G	0.0	2.1	10.8	9.2
PCR_RCP26_H	0.3	2.3	18.3	10.7
PCR_RCP26_I	0.0	2.7	17.0	13.3
PCR_RCP26_M	0.1	2.4	17.5	12.5
WAT_RCP26_G	0.0	2.1	9.5	7.1
WAT_RCP26_H	0.4	2.2	13.9	7.4
WAT_RCP26_I	0.3	2.4	14.1	9.7
WAT_RCP26_M	0.3	2.5	16.3	7.1
ORC_RCP26_G	-0.3	2.3	8.7	13.6
ORC_RCP26_I	-0.5	2.9	11.0	20.3

Table S5-6. Characteristics of projected trend (GCMRCP60) across 18 members at the global scale. Mean and standard deviation have unit of %-change per decade. Cell-based significance was identified using the Mann-Kendall test at the 10% significance level. Note that no statistical test for comparison was conducted.

Streamflow simulations	Trend mean	Trend standard deviation	Percentages of significant	
			Increasing trend	Decreasing trend
H08_RCP60_G	0.6	3.4	22.2	16.4
H08_RCP60_H	1.0	4.4	29.8	17.3
H08_RCP60_I	-0.1	3.7	18.2	26
H08_RCP60_M	0.7	3.7	28.7	14.3
LPJ_RCP60_G	0.0	2.6	18.5	15.2
LPJ_RCP60_H	-0.1	3.5	24.5	21.3
LPJ_RCP60_I	-0.5	3.2	16.3	23.9
LPJ_RCP60_M	0.2	3.0	23.5	15.7
PCR_RCP60_G	0.0	3.0	20.0	18.0
PCR_RCP60_H	0.2	3.8	28.2	21.4
PCR_RCP60_I	-0.4	3.6	20.5	24.5
PCR_RCP60_M	0.5	3.0	29.4	13.8
WAT_RCP60_G	0.5	2.7	24.9	9.5
WAT_RCP60_H	0.9	3.4	31.3	10.4
WAT_RCP60_I	0.2	3.4	23	16.3
WAT_RCP60_M	0.9	3.2	31.2	9.3
ORC_RCP60_G	-0.2	3.0	17.6	20.2
ORC_RCP60_I	-1.1	4.1	14.7	33.6

Chapter 6. Conclusions

Improved understanding of trends in flood hazards and the driving mechanisms of detectable trends at the global scale are of scientific and social importance. Although numerous scientific efforts have been invested into detecting and attributing trends in flood hazards, a consistent picture at the global scale is not yet available due to several barriers including (i) limited spatiotemporal coverage of streamflow observations, (ii) fragmented (and even conflicting) findings of regional scientific studies, and (iii) the complexity of flood generating processes that make it difficult to upscale conclusions at the local or regional scale to the global scale, or translate findings from one region to another. With global availability and extensive temporal coverage, streamflow simulations from global hydrological models (GHMs) have been increasingly used as an alternative to assess changes in flood hazards and attribute these patterns to climate and catchment processes at the global scale. The trade-off, however, is the uncertain quality of simulated trends of GHMs, mainly due to the acknowledged high uncertainty in model structure, inputs data, and model parameterisation. Owing to these challenges, a coherent picture of how and why flood hazards have changed at the global scale has not yet been available.

This thesis has provided an observation-based perspective of floods hazards to address the presented challenges. The objectives of this thesis were:

1. **Objective 1:** to improve the spatiotemporal coverage of global streamflow observation system.
2. **Objective 2:** to complement the understanding of changes in flood hazard and the driving mechanisms at the global scale.
3. **Objective 3:** to evaluate the capacity of GHMs in reproducing changes in flood hazards.

The research within this thesis is comprehensive, taking advantage of the available global streamflow databases together with the latest global data

products such as climate reanalysis and streamflow simulations from GHMs. This chapter describes the main contributions of the research, outlines the key limitations that have not yet been addressed, and suggests potential work to further improve the understanding of flood hazards at the global scale.

6.1 Research contribution

This research comprises several large-sample investigations that have yielded important findings on the characteristics of floods at the global and continental scales. The release of the Global Streamflow Indices and Metadata (GSIM) archive, developed as part of this thesis, enables new opportunities for advances in hydrological research through better spatiotemporal observation coverage. The thesis therefore represents a significant contribution to improved understanding of flood characteristics at the global scale. The specific research contributions are outlined below:

1. **The publication of an open database of daily streamflow observations over more than 30,000 gauges (Objective 1).** The Global Streamflow Indices and Metadata (GSIM) archive has improved spatiotemporal coverage, including comprehensive metadata including such as the catchment boundary and a range of landscape attributes for each gauging station. The volume of data is unprecedented, representing a four-fold increase in data availability relative to the next-largest daily streamflow database currently available (i.e. the Global Runoff Data Base, the core database of GSIM). It demonstrates the possibility of improving the observation system through the collation of databases that are freely accessible online. The data products (Supplementary Dataset 1 and Supplementary Dataset 2), which are publicly available on the PANGEAE data repository, provide an enhanced opportunity to explore changes in streamflow regimes.

2. **An improved understanding of historical changes in flood hazards at the global scale (Objective 2).** Chapter 3 investigated observed trends in flood hazards and the impacts of catchment characteristics on changes in floods. Previous studies of flood trends were fragmented and focused on specific regions or basins, with different study parameters such as the time

period and length of record. This paper was the first to provide analysis at a global scale with consistency in the method of analysis using a large sample of catchments. The research found the substantial importance of catchment size while there was no significant evidence for the importance of other factors such as forest change, climate condition (represented by Koppen climate classes) and dam presence.

3. **A global-scale prediction of flood seasonality built on a comprehensive analysis of dominant flood generation mechanisms (Objective 2).** The investigation of flood seasonality, presented in Chapter 4, identifies the dominant flood generation processes from continental to global scale. Homogeneous regions in terms of the most important flood generation processes were generalised across the globe using climate indices derived from an atmospheric reanalysis product to enable global prediction for flood timing. The framework presented in Chapter 4 is useful for assessing the implications of climate change and variability on the timing of flood occurrence.

4. **An evaluation of GHM capacity in reproducing trends in flood hazards (Objective 3).** The model-observation comparison of trends in flood hazard presented in Chapter 5 represents the first attempt to evaluate GHM capacity in terms of reproducing trends in flood hazards. The assessment across 3666 locations globally shows moderate capacity of GHMs in simulating the mean and pattern of observed trends in flood hazard, while the performance is limited in terms of the spread of trends. The comparison also highlights the importance of climate input uncertainty on GHMs, as simulations forced with modelled climate inputs have significantly lower performance compared to simulations forced with observational climate data.

5. **An assessment of projected changes in flood hazard under two socio-economic scenarios (Objective 2).** The assessment of projected trends in streamflow extremes presented in Chapter 5 complements current state of understanding about the implication of climate model uncertainty in developing projections of flood hazards. Under a 'business-as-usual' scenario (GCMRCP60), some regions, e.g. south-eastern Asia, eastern Africa, Siberia and Greenland, were consistently projected with significant

increasing trend. These 'high-risk' regions, however, are only covered by less than 1% of all GSIM stream-gauges, indicating a highly uncertain future for flood-prone communities in a warming climate.

The Supplementary Articles and Supplementary Datasets of this thesis are also significant contributions to improved understanding of flood hazards at the global scale. Supplementary Article 1 and the associated dataset (Supplementary Dataset 2) described the production of a comprehensive set of streamflow indices at monthly/seasonal/yearly resolutions, capturing the important aspects of streamflow regimes (e.g. quantiles of streamflow, centre timing, the timing of low-/high-flow) across more than 30,000 locations globally. Supplementary Article 2 reported the to-date most comprehensive analysis of indicators for mean and extremes streamflow indices. The regional assessment within this study suggested unidirectional changes of the whole flow distribution towards either wetter or drier conditions, indicating the rise in flood hazard may be compensated by reduced drought hazard or vice versa. Supplementary Article 3 reviewed the progress and challenges in producing large-sample datasets, the backbone for large-sample hydrology investigations, and provides guidelines towards better coverage for large-sample datasets. Together, the supplementary materials combined with the main chapters of the thesis serve the global hydrology community with a better understanding of flood hazard characteristics from an observation-based perspective.

6.2 Limitations and future work

Although this research has contributed a comprehensive observation-based perspective on global flood hazards, there are some challenges that have not been fully addressed within the context of this thesis. The following sections summarise the limitations and outline opportunities to address them.

Data coverage remains a challenge to yield an observation-based conclusion about changes in flood hazard over some flood-prone regions. Although GSIM has unprecedented coverage, there is scope to further advance the coverage. The limited number of component databases (12 in total) of GSIM means that improvement was only available for some countries such as India, Japan, Brazil, and Russia. Many countries in flood-prone regions, e.g. the East

and South Asia countries, are still sparsely represented. When relatively strict data entrance criteria are applied (e.g. studies in Chapter 4 and Chapter 5), data over these countries are completely removed. The data-related challenges can only be addressed with the efforts from many stakeholders: data providers (for making more streamflow records accessible), peer-reviewed publishers (for requiring data that was used in published studies available in the public domain), and relevant international organisations (such as WMO or GRDC; to endorse and support efforts in streamflow data exchange).

The need to break down the importance of natural and anthropogenic factors driving floods at the continental and global scale. The global prediction of flood timing (Chapter 4), which uses only the most important flood generation process to predict flood seasonality, has poor performance in several regions. This result indicates high spatial variability in terms of flood generation processes at the sub-continental level, with more than one mechanism contributing substantially to flood production at many locations. The analysis of the impacts of catchment characteristics on changes in flood hazard (Chapter 3) suggests the potential influence of non-climatic factors (e.g. catchment size, topography, river morphology, land use change) to flood hazard characteristics (e.g. magnitude, frequency or timing), which is also a significant factor to be further explored in flood studies at the global scale. One possibility to explore this research pathway is through the use of the recent data products from related fields. For instance, GHMs (with global coverage and the possibility to switch on and off human influence) are able to further investigate the implications of climate change and human water management on flood hazard characteristics. The recent publication of sub-daily precipitation products with global coverage (Sun et al., 2018) provides an opportunity to quantify the importance of rainfall bursts to flood occurrence. Advanced land-cover related data-products (e.g. land-use change, forest cover change, the evolution of irrigation area) enable the possibility to sensitivity-test the influence of human activity.

A consistent picture of changes in global flood hazard between observation-based and modelling-based studies remains elusive. There has been increasing evidence of substantial impacts of both anthropogenic climate

change as well as human water and land management on the world's freshwater resources and hydrological extremes. A global picture of changes in flood hazards, however, has not yet been available due to imperfect model estimates and limited spatiotemporal coverage of relevant in-situ observations. The comparison between observed trends and simulated trends using GHMs indicates moderate model performance, but still shows limited quality of simulated trends over some areas, especially Australia and Asia. Uncertainty in climate models, which were found to have poor performance in reproducing observed trends for climate indices across many Southern Hemisphere regions, also reduces the credibility of GHM simulations forced with modelled climate data. Advances in flood hazard understanding at the global scale, therefore, requires combined effort from multiple communities. Besides the important role of data agencies in improving observational streamflow coverage (discussed above), the modelling community also needs to tackle long-standing challenges in simulating global water resources such as improving the parameterisations of human influence and catchment processes, and downscaling GHMs to finer spatial resolution while maintaining correct model structures/concepts. The uncertainty in climate models should also be accounted for, e.g. by using observations to constrain climate simulations (Padrón et al., 2019, He and Soden, 2016), to provide a more reliable modelled atmospheric variables (e.g. precipitation and temperature) that are arguably the most important factor determining the ability of the large-scale hydrological models.

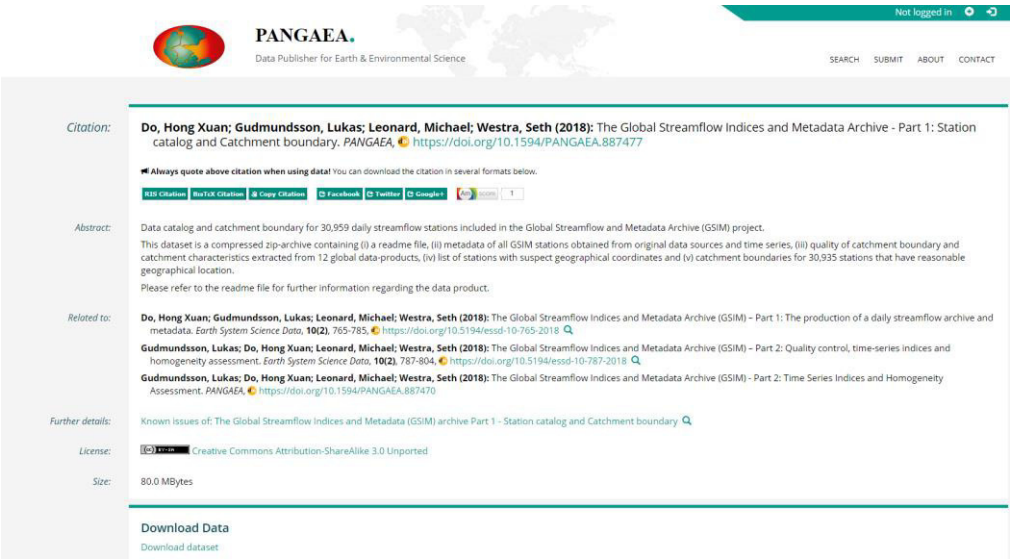
Addressing these research challenges requires further improvements to the spatiotemporal coverage of streamflow observations across the global land mass. It is therefore recommended more data to be contributed to international streamflow archives such as the Global Runoff Data Base or GSIM to further improve accessible in-situ streamflow records to the broad hydrologic community. GHMs, with the possibility to hypothesis-test the implications of varying natural/anthropogenic scenarios on flood characteristics, remain a crucial toolset to complement observation-based findings of flood hazard characteristics. As a result, large-sample evaluations of GHM performance (Veldkamp et al., 2018; Zaherpour et al., 2018; Zhao et al., 2017) exploiting

global databases such as the Global Runoff Data Base or GSIM (Do et al., 2018a, b; Gudmundsson et al., 2018a, b) should also be emphasised to further quantify the reliability of GHMs in characterising the many aspects of streamflow regimes at the global scale.

Supplementary publications

Supplementary Dataset 1

Do, H. X.; Gudmundsson, L., Leonard, M., Westra, S. (2018) The Global Streamflow Indices and Metadata Archive - Part 1: Station catalog and Catchment boundary. PANGAEA, https://doi.org/10.1594/PANGAEA.887477



Supplementary Dataset 2

Gudmundsson, L., Do, H.X., Leonard, M., Westra, S. (2018) The Global Streamflow Indices and Metadata Archive (GSIM) - Part 2: Time Series Indices and Homogeneity Assessment. PANGAEA, https://doi.org/10.1594/PANGAEA.887470

Supplementary Article 1

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S. (2018) The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787-804, https://doi.org/10.5194/essd-10-787-2018.

Earth Syst. Sci. Data, 10, 787–804, 2018
https://doi.org/10.5194/essd-10-787-2018

Open Access Earth System Science Data

The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment

Lukas Gudmundsson[1], Hong Xuan Do[2], Michael Leonard[2], and Seth Westra[2]

[1]Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH Zurich, Universitaetstrasse 16, Zurich 8092, Switzerland
[2]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, Australia

Correspondence: Lukas Gudmundsson (lukas.gudmundsson@env.ethz.ch)

Received: 7 September 2017 – Discussion started: 20 September 2017
Revised: 9 March 2018 – Accepted: 16 March 2018 – Published: 17 April 2018

Abstract. This is Part 2 of a two-paper series presenting the Global Streamflow Indices and Metadata Archive (GSIM), which is a collection of daily streamflow observations at more than 30 000 stations around the world. While Part 1 (Do et al., 2018a) describes the data collection process as well as the generation of auxiliary catchment data (e.g. catchment boundary, land cover, mean climate), Part 2 introduces a set of quality controlled time-series indices representing (i) the water balance, (ii) the seasonal cycle, (iii) low flows and (iv) floods. To this end we first consider the quality of individual daily records using a combination of quality flags from data providers and automated screening methods. Subsequently, streamflow time-series indices are computed for yearly, seasonal and monthly resolution. The paper provides a generalized assessment of the homogeneity of all generated streamflow time-series indices, which can be used to select time series that are suitable for a specific task. The newly generated global set of streamflow time-series indices is made freely available with an digital object identifier at https://doi.pangaea.de/10.1594/PANGAEA.887470 and is expected to foster global freshwater research, by acting as a ground truth for model validation or as a basis for assessing the role of human impacts on the terrestrial water cycle. It is hoped that a renewed interest in streamflow data at the global scale will foster efforts in the systematic assessment of data quality and provide momentum to overcome administrative barriers that lead to inconsistencies in global collections of relevant hydrological observations.

1 Introduction

Although terrestrial freshwater is an essential component of the Earth system and a prerequisite for societal development, the availability of relevant in situ observations at the global scale has been limited. Until now, most relevant in situ observations have been held by national and regional authorities, and despite their best efforts, international data centres only have access to a small subset of the full observed record (Do et al., 2018a). This situation stands in contrast to the fact that monitoring data are increasingly being made publicly available through regional and national authorities (Do et al., 2018a). In this paper series, we present an international collection of river and streamflow observations that covers more than 30 000 stations around the globe, highlighting the fact that these are among the best monitored variables of the terrestrial water cycle (Fekete et al., 2012, 2015; Gudmundsson and Seneviratne, 2015; Hannah et al., 2011). Part 1 of the paper series (Do et al., 2018a) documents the data-collection process together with a meta-database that allows users to recreate the collection from the original data sources. In addition, Part 1 of this paper series also presents auxiliary data including catchment boundaries delineated from global digital elevation models as well as selected properties (e.g. land cover, climate) of these catchments.

While the data collection outlined in Part 1 (Do et al., 2018a) increases the spatial and temporal availability of streamflow records at the global scale, it is important to also

consider the quality of the data. This is especially relevant for this merged data product combining information from several databases, which might have been set up with different objectives. Furthermore, data contained in individual databases may stem from different sources, often with unknown quality control procedures. In addition, changes in instrumentation as well as human impacts such as stream straightening or flow regulations can have pronounced effects on the observed record. Establishing a database of quality controlled streamflow observations is therefore essential for many applications, including e.g. the need to evaluate the increasing number of continental- and global-scale hydrological and land-surface models that have emerged in recent decades (Beck et al., 2017; Gudmundsson et al., 2012a, b; Haddeland et al., 2011; Zaitchik et al., 2010) and the assessment of human impacts on the terrestrial water cycle (Alkama et al., 2013; Barnett et al., 2008; Destouni et al., 2013; Gudmundsson et al., 2017; Hegerl et al., 2015; Hidalgo et al., 2009; Jaramillo and Destouni, 2015; Oliveira et al., 2011). While there have been significant efforts in the climatological community to share and standardize transnational weather observations as well as derivative data products (Alexander et al., 2006; Becker et al., 2013; Dee et al., 2011; Harris et al., 2014; Haylock et al., 2008; Poli et al., 2016), the hydrological community has traditionally been reticent to adopt regional or global approaches, instead focussing predominantly on the catchment scale. A more concerted and coordinated effort to understand the quality of streamflow observations across the globe provides significant opportunities for fostering hydrological research in support of understanding of global water budgets. This paper initiates the process of evaluating, analysing and documenting the quality of observed streamflow time series, providing a method for increasing the reliability and ongoing value of the database. To do so, this paper expands on previous research (Gudmundsson and Seneviratne, 2016) and applies a set of transparent and reproducible methods to evaluate the quality of the considered records.

One limitation of the newly assembled collection of daily river flow and streamflow time series is that publication of unprocessed daily values is restricted for some of the original data sources. To nevertheless be able to publish relevant information on observational streamflow, we therefore present here processed data in the form of time-series indices that capture essential aspects of (i) the water balance, (ii) seasonality, (iii) low flows and (iv) floods. The approach of publishing time-series indices instead of raw daily values is adapted from the CCl/WCRP/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) (https://www.wcrp-climate.org/data-etccdi), which has developed this approach to make relevant climate information publicly available in cases where access to raw daily values is restricted. The ETCCDI has focussed on indices characterizing changes in extreme precipitation and temperature, based on a core collection of indices proposed by Frich et al. (2002). Both Klein Tank et al. (2009) and Zhang et al. (2011) provide additional background on the usage and computation of the ETCCDI indices. Klein Tank et al. (2009) also provide guidelines for quality control of the raw daily input data, index computation and assessment of time-series homogeneity.

The use of time-series indices for characterizing the temporal evolution of selected river flow characteristics is also common practice in the hydrological literature. Typically used time-series indices include mean annual flows (e.g. Kumar et al., 2009; Milly et al., 2005; Small et al., 2006; Stahl et al., 2010, 2012), indices that can be used to characterize changes in the seasonal cycle (e.g. Blöschl et al., 2017; Cunderlik and Ouarda, 2009; Ehsanzadeh and Adamowski, 2010; Hidalgo et al., 2009; Moore et al., 2007; Rauscher et al., 2008; Regonda et al., 2005; Stewart et al., 2005), time series of annual percentiles (e.g. Gudmundsson et al., 2011; Lins and Slack, 1999; Zhang et al., 2001), flood indices (e.g. Blöschl et al., 2017; Hodgkins et al., 2017; Kumar et al., 2009; Kundzewicz et al., 2005; Lins and Slack, 1999; McCabe and Wolock, 2002; Small et al., 2006; Svensson et al., 2005; Zhang et al., 2001) and low-flow indicators (e.g. Hisdal et al., 2001; Lins and Slack, 1999; McCabe and Wolock, 2002; Small et al., 2006; Stahl et al., 2010, 2012; Svensson et al., 2005; Tallaksen et al., 1997; Zhang et al., 2001).

In addition, several studies have focussed on collections of hydrological signatures (or flow characteristics) that are designed to summarize long-term properties of observed river flow and streamflow (e.g. 2013; Beck et al., 2015; Olden and Poff, 2003; Sawicz et al., 2011, 2014; Westerberg et al., 2016). These hydrological signatures include e.g. mean annual flow, flow percentiles, characteristics of the flow duration curves, indications of seasonality and the base flow index. These signatures are typically derived from all daily values in a long time window (e.g. the base flow index computed from all daily values from 1985 to 2010). This is an important structural difference if compared to time-series indices, which are typically computed every year, every season or every month (e.g. time series of annual maxima) and thus also allow for an assessment of changing hydrological conditions over time.

The following sections build upon these efforts and present a collection of quality controlled river and streamflow time-series indices. To do so, we first introduce an approach to check the quality of individual daily observations using a combination of information provided with the original data and data-driven procedures. Subsequently we present a collection of time-series indices that can be computed for yearly, seasonal and monthly resolution. An assessment of the statistical homogeneity of the newly derived indices is provided to allow users to filter the published data according to their own eligibility criteria. Given that each application may warrant a different assessment of the trade-off between the quantity and quality of available data, the presented collection of streamflow time-series indices has sought to avoid predefined eligibility criteria (such as predefining a base period

Table 1. Quality flags of daily values of all databases that enter the GSIM collection (see Do et al., 2018a).

Database	Quality code
GRDB	Not recommended by data provider. There are four flags: −999 – missing data, no correction 1 – corrected data, no method specified 99 – usage not recommended by the provider 900 – calculated from daily water level Note: in recent updates GRDC does not provide quality flags.
EWA	Not recommended by data provider (similar to GRDB)
ARCTICNET	Quality flag not provided
GAME	Quality flag not provided
CHDP	Quality flag not provided
USGS	Flags were provided for each data point. There are four categories: A: value has been validated to be published A:e: value was estimated and validated to be published P and P:e: Provisional data
BOM	Flags were provided for each data point. There are five categories documented: A (flag 10): best available data B (flag 90): compromised to represent the parameter C (flag 110): estimated value E (flag 140): quality is not known F (flag 210): poor quality or missing Flag "−1" also presents to indicate missing value
HYDAT	Quality flags were only provided for some data points. There are five categories documented: A: Partial Day (numeric value 1) B: Ice Conditions (numeric value 2) D: Dry (numeric value 3) E: Estimated (numeric value 4) S: Sample(s) collected this day (numeric value 5)
WRIS	Qc flag not provided
ANA	Quality flags were only provided for some data points. Flags were described in Portuguese. Below are the authors' interpretation. 0: No description 1: Real measurement 2: Estimated 3: Doubtful 4: Dry
MLIT	Quality flag not provided
AFD	Quality flag not provided

or presupposing only high-quality sites). The paper closes with an open invitation to the hydrological and Earth science communities on how to best facilitate activities that might lead to sustained collation, curation and improvement of global streamflow data.

2 Quality control (QC) of daily values

2.1 Strategy for QC of daily values

As the considered data stem from several sources, some of which have a complex history, it is difficult to a priori judge the quality of individual records. Ideally, each of the considered series would be accompanied by detailed information on the station properties (e.g. information on sensors or the design of the gauging weir) and on the credibility of indi-

Table 2. Translation of daily quality control (QC) flags of the original databases (Table 1) to standardized values prior to the calculation of indices. Note that the Global Runoff Data Centre advises not to consider the QC flags in the GRDB and EWA files. Note also that some databases (HYDAT, ANA) do not provide QC flags for all daily data.

Dataset	QC flag not provided	QC flags are not recommended by data provider	Reliable	Suspect
GRDB	–	All data points	–	–
EWA	–	All data points	–	–
ARCTICNET	All data points	–	–	–
GAME	All data points	–	–	–
CHDP	All data points	–	–	–
USGS	–	–	"A" and "A:e" (approved data)	"P" and "P:e" data (provisional data)
BOM	–	–	A (table below)	B, C, D, F (table below)
HYDAT	Other data points	–	B, D, S	A, E
WRIS	All data points	–	–	–
ANA	0, no value	–	1, 4	2, 3
MLIT	All data points	–	–	–
AFD	All data points	–	–	–

vidual daily values. However, this information is often not available or difficult to access and only some of the original data sources provide daily quality flags (Table 1). In addition, the large number of languages involved and the sheer quantity of gauging stations render a detailed manual assessment unfeasible. Nevertheless, it is essential to apprise the quality of individual observations prior to any assessment. As some of the considered time series come with daily quality flags (usually based on simple plausibility checks), while others do not, the two cases are treated separately.

2.2 Quality control of daily values if reliable flags are provided

As noted in Do et al. (2018a), some of the considered databases provide quality control (QC) flags for daily values that distinguish between reliable and suspect observations (Table 1). To allow for a combined assessment, the original QC flags were translated into a common set that distinguishes *suspect* from *reliable* values (Table 2). This step is necessary for consistency, since some databases provide a variety of QC flags to indicate suspect cases, but neither the same flags nor the level of fidelity are available across all databases. Regarding the Global Runoff Data Centre (GRDC), while QC flags are available in the EWA and GRDB files entering the presented collection, the GRDC advised not to use them. In these cases, the time series are treated as if no QC flags were provided. Note also that the GRDC has discontinued QC flagging in the latest version of the data. Some databases do not provide QC flags for every time step (Table 2); in these cases time steps without original QC flags were assumed to be *reliable* as long as at least one time step was flagged in the respective time series.

2.3 Quality control of daily values if no reliable flags are available

For original time-series files for which no QC flags are available or for which there is advice against using available QC flags by the data providers (GRDB and EWA), automated techniques can be used to classify the reliability of individual daily data points using simple and reproducible tests focussing on the plausibility of individual values. The following three criteria are based on a previously used procedure (Gudmundsson and Seneviratne, 2016), were developed on the basis of techniques described in Reek et al. (1992) and the ECA & D Project Team and Royal Netherlands Meteorological Institute (2013; later referred to as EAC&D13), and were further refined using suggestions on outlier detection for index calculation by Klein Tank et al. (2009):

1. Days for which $Q < 0$ are flagged as *suspect*, where Q denotes a daily streamflow value. The rationale underlying this rule is that streamflow values smaller than zero are non-physical (Gudmundsson and Seneviratne, 2016).

2. Daily values with more than 10 consecutive equal values larger than zero are flagged as *suspect*. This rule is motivated by the fact that many days with consecutive streamflow values often occur due to instrument failure (e.g. damaged sensors, ice jams) or flow regulations. The threshold of 10 days is a compromise chosen to account for the possibility that consecutive equal observations may reflect the truth e.g. if day-to-day fluctuations are below the sensitivity of the employed sensor (Gudmundsson and Seneviratne, 2016).

3. Based on a previously suggested approach for evaluating temperature series (Klein Tank et al., 2009), daily streamflow values are declared as outliers if values of $\log(Q + 0.01)$ are larger or smaller than the mean value

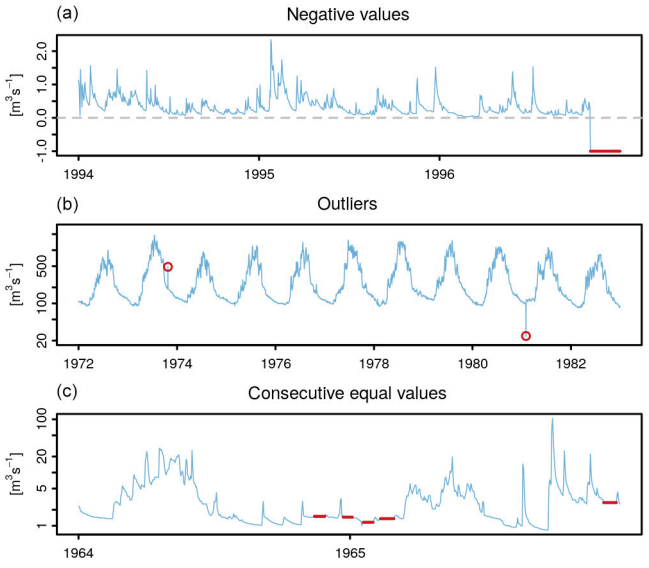


Figure 1. Three example time series illustrating issues detected by the three daily quality control criteria (highlighted in red). The first panel shows negative values at the end of the time series of Rohr at Rohrhardsberg, Germany. The second panel shows two outliers detected in the time series of Vakhsh at Gram, Tajikistan. The third panel shows instances of more than 10 consecutive equal values found in the time series of Tanara at Ponte di Nava, Italy. Note that all time series were trimmed for visualization purposes. Note also the logarithmic axis in panels two and three.

of $\log(Q+0.01)$ plus or minus 6 times the standard deviation of $\log(Q+0.01)$ computed for that calendar day for the entire length of the series. The mean and standard deviation are computed for a 5-day window centred on the calendar day to ensure that a sufficient amount of data is considered. The log-transformation is used to account for the skewness of the distribution of daily streamflow values and 0.01 was added because the logarithm of zero is undefined. Outliers are flagged as *suspect*. The rationale underlying this rule is that unusually large or small values are often associated with observational issues. The 6 standard-deviation threshold is a compromise, aiming at screening out outliers that could come from instrument malfunction, while not flagging extreme floods or low flows.

An example of the outcome of this automated quality control of daily observations is shown in Fig. 1, which displays daily streamflow observations at three locations and highlights time steps that did not pass the three above-mentioned criteria. Note that the outlier detection (middle panel) did not screen out extreme floods or low flows, but only values that were unusually large or small for the respective time of the year, where one case involves a spurious large flow and the other a spurious small flow.

3 Streamflow indices

3.1 General considerations, design rules and reliability

3.1.1 General considerations

Table 3 describes a set of streamflow time-series indices that are designed to facilitate the analysis of (i) changes in the regional water balance, (ii) changes in the seasonal cycle, (iii) floods, and (iv) low flows. Many of the considered indices have been previously used in the scientific literature and Table 4 presents, wherever possible, a selection of relevant references and additional information. Note also that index selection was limited to those that can be computed without a base period, which excludes many; examples include "the number of days in a year, or season, for which daily values exceed a time-of-year-dependent threshold" (Zhang et al., 2005), drought deficit volumes (Loon and Anne, 2015; Tallaksen et al., 1997) and anomalies with respect to a climatological normal (McKee et al., 1993; Shukla and Wood, 2008). There are two reasons for excluding these indices: first, regional differences in temporal coverage hinder an unambiguous identification of a common base period that can be used around the globe. Second, it is now well established that indices that depend on a base period are prone to inhomogeneities if the base period is shorter than the considered series (Sippel et al., 2015; Zhang et al., 2005). Although both analytical (Sippel et al., 2015) and non-parametric (Zhang et al., 2005) solutions exist to mitigate this problem, we chose not to include indices that require a base period. This is because the available solutions either depend on strong normality assumptions (Sippel et al., 2015) or are computationally intensive (Zhang et al., 2005), which implies that the time-series indices cannot be easily extended when new data become available. Finally, it is noteworthy to mention that indices are easier to update when they do not have a base period, as they can be computed without knowledge of previous values.

3.1.2 Design rules for index calculation

The design rules for calculating time-series indices closely follow the recommendations of ECA&D13. Before index calculation, all daily values that are flagged as *suspect* by the daily QC procedure are set to missing, and indices are computed using the remaining data points. All indices are computed on yearly time steps, while some indices are also computed with seasonal and monthly resolution. Seasons are defined as December–January–February (DJF), March–April–May (MAM), June–July–August (JJA) and September–October–November (SON). The reason for not computing all indices for seasonal and monthly resolutions is related either to the fact that some indices are only defined on annual timescales, or to the amount of data required for reliable computation. All considered indices are described in Tables 3 and 4.

Table 3. Definition of time-series indices contributing to the GSIM archive. Abbrev. Indicates the abbreviation of the index name used throughout this paper as well as in the database. Resol. indicates the time resolution for which the index is computed, which can take values of Y (yearly), seasonal (S) and monthly (M).

Title	Abbrev.	Units	Resol.	Definition
Mean daily streamflow	MEAN	($m^3\ s^{-1}$)	Y, S, M	Arithmetic mean of daily streamflow.
Standard deviation of daily streamflow	SD	($m^3\ s^{-1}$)	Y, S, M	Standard deviation of daily streamflow.
Coefficient of variation of daily streamflow	CV	(–)	Y, S, M	Standard deviation of daily streamflow divided by the mean daily streamflow (SD/MEAN).
Interquartile range of daily streamflow	IQR	($m^3\ s^{-1}$)	Y, S, M	75th–25th percentile of daily streamflow.
Minimum daily streamflow	MIN	($m^3\ s^{-1}$)	Y, S, M	Minimum value of daily streamflow.
Maximum daily streamflow	MAX	($m^3\ s^{-1}$)	Y, S, M	Maximum value of daily streamflow.
Minimum 7-day mean streamflow	MIN7	($m^3\ s^{-1}$)	Y, S, M	Minimum 7-day arithmetic mean streamflow. For computation, the complete daily time series are first smoothed with a backward looking moving average with a 7-day window. Subsequently, the minimum value for each yearly, seasonal or monthly period is determined.
Maximum 7-day mean streamflow	MAX7	($m^3\ s^{-1}$)	Y, S, M	Maximum 7-day arithmetic mean streamflow. For computation, the complete daily time series are first smoothed with a backward looking moving average with a 7-day window. Subsequently, the maximum value for each yearly, seasonal or monthly period is determined.
10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th percentiles of daily streamflow	P10, P20, P30, P40, P50, P60, P70, P80, P90	($m^3\ s^{-1}$)	Y, S	Percentile values of daily streamflow computed for each yearly and seasonal period, where low percentiles (e.g. 10th percentile) correspond to low flows.
Centre timing	CT	(doy)	Y	The day of the year (doy) at which 50 % of the annual flow is reached. The index is computed for calendar years, where 1 denotes 1 January.
Day of minimum streamflow	DOYMIN	(doy)	Y	The day of the year (doy) at which the minimum flow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years.
Day of maximum streamflow	DOYMAX	(doy)	Y	The day of the year (doy) at which the maximum flow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years.
Day of minimum 7-day mean streamflow	DOYMIN7	(doy)	Y	Day of the year (doy) at which the minimum 7-day arithmetic mean streamflow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years. For computation, the daily time series is first smoothed using a backward looking moving average with a 7-day window length. Subsequently, the day of the minimum of each year is determined.
Day of maximum 7-day mean streamflow	DOYMAX7	(doy)	Y	Day of the year (doy) at which the maximum 7-day arithmetic mean streamflow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years. For computation, the daily time series is first smoothed using a backward looking moving average with a 7-day window length. Subsequently, the Julian day of the maximum of each year is determined.
Gini coefficient	GINI	(–)	Y	For daily runoff values q of each year, that are sorted with index i in increasing order such that $q_i \leq q_{i+1}$ GINI is defined as $\frac{1}{n}\left(n+1-2\left(\frac{\sum_{i=1}^{n}(n+1-1)q_i}{\sum_{i=1}^{n}q_i}\right)\right)$, where n is the number data points available for that year. The Gini coefficient ranges from 0 to 1. Values of 0 indicate uniform distribution of flows throughout the time period (i.e. year), whereas values close to 1 indicate that all the flows occur on a single day.

Table 4. Commentary and literature supporting the GSIM indices.

Abbrev.	Commentary
MEAN	Mean daily streamflow is a commonly used water-balance measure and often used as a proxy for renewable freshwater resources (Oki and Kanae, 2006; Shiklomanov et al., 2004; Vörösmarty et al., 2000). Observed time series of mean yearly or monthly streamflow has e.g. been subject to trend analysis at regional to continental scales (e.g. Kumar et al., 2009; Lettenmaier et al., 1994; Lins and Slack, 1999; Milly et al., 2005; Small et al., 2006; Stahl et al., 2010, 2012).
SD	The standard deviation of daily streamflow provides information on the total variability for each yearly, seasonal and monthly time step. This index therefore includes information related to floods and low flows as well as the amplitude of the annual cycle (yearly only). We are not aware of any study analysing time series of the standard deviation of daily streamflow.
CV	The coefficient of variation of daily streamflow is a relative measure of daily variability. In contrast to SD, CV is independent of the mean flow and does hence allow for an isolated assessment of day-to-day streamflow variability. We are not aware of any study analysing time series of the coefficient of variation of daily streamflow.
IQR	The inter quartile range is a measure of day-to-day streamflow variability. Through its definition as the difference between the 75th and 25th percentiles, the IQR provides information on the width of the centre of the distribution and is less sensitive to extreme outliers than SD or CV. We are not aware of any study analysing time series of the standard deviation of daily streamflow.
MIN	Minimum daily streamflow is a regularly used low-flow indicator. Especially the yearly minimum has been used widely as it is an easy to interpret measure and lends itself to analysis in the framework of the generalized extreme value distribution (Tallaksen and van Lanen, 2004). Annual minimum streamflow series are also commonly subject to large-scale trend analysis (Kumar et al., 2009; Lins and Slack, 1999; McCabe and Wolock, 2002; Zhang et al., 2001).
MAX	Maximum daily streamflow is a widely used indicator for high flows and floods. Especially annual maximum time series are regularly considered as they allow for a straightforward interpretation and can easily be analysed through the generalized extreme value distribution (Katz et al., 2002). Time series of annual maximum streamflow have been subject to regional and global trend assessments (e.g. Do et al., 2017; Hall et al., 2015; Kumar et al., 2009; Kundzewicz et al., 2005; Lins and Slack, 1999; McCabe and Wolock, 2002; Small et al., 2006; Zhang et al., 2001).
MIN7	Time series of minimum 7-day mean streamflow have been repeatedly used as a low-flow and drought metric. Through the smoothing operation, MIN7 is less sensitive to small day-to-day fluctuations, but focusses on sustained periods with limited water availability. MIN7 time series have e.g. been subject to large scale trend assessments (Kumar et al., 2009; Small et al., 2006; Stahl et al., 2010; Svensson et al., 2005).
MAX7	Time series of 7-day mean maximum streamflow do not focus on the highest water levels ever recorded, but rather on sustained periods of very high flow. Time series of MAX7 have e.g. been used to assess streamflow trends in India (Kumar et al., 2009; Stahl et al., 2012).
P10, P20, P30, P40, P50, P60, P70, P80, P90	Percentiles of daily streamflow provide together with MIN and MAX an approximation of the empirical cumulative distribution function (ECDF) of daily streamflow for each considered seasonal or yearly time period. These indices are not provided on monthly resolution, as it appears to be excessive to compute percentiles in 10 % steps based on 28 to 31 daily values. Note also that an alternative definition of the ECDF is also referred to as the flow-duration curve (FDC) in the hydrological literature. The difference between the ECDF and the FDC is that the FDC uses an inverse definition of percentiles (exceedance frequencies), such that high values correspond to low flows (Tallaksen and van Lanen, 2004; Vogel and Fennessey, 1994). Besides approximations of the ECDF, the percentile series can be used to characterize "moderate extremes" (Zhang et al., 2011), i.e. very high or very low values that can occur several times each year and are hence more robust to quantify. Sets of annual percentile series have for example been used to investigate regional low- and high-flow dynamics in Europe (Gudmundsson et al., 2011) and have been subject to regional-scale trend assessments (Lins and Slack, 1999; McCabe and Wolock, 2002; Zhang et al., 2001).
CT	The centre timing is an index that is sensitive to changes in the seasonal cycle. Lower values indicate that more than half of the annual discharge has occurred earlier in the year. That means, that values smaller or equal than 182 would correspond to a year for with at least half of the streamflow volume has occurred in the first half of the year. Note that CT is usually defined for hydrological years in the literature and that the precise definition of CT can vary between studies (Hidalgo et al., 2009; Moore et al., 2007; Rauscher et al., 2008; Regonda et al., 2005; Stewart et al., 2005). Here we compute CT for calendar years to ensure consistency with the remaining indices and because the definition of the hydrological year depends on local climate conditions. Time series of CT have been used to assess changes in the timing of the seasonal cycle of streamflow in several regional studies (Hidalgo et al., 2009; Moore et al., 2007; Rauscher et al., 2008; Regonda et al., 2005; Stewart et al., 2005).
DOYMIN	The timing of annual minimum flow can provide valuable information on the processes underlying low flows. For example, in snowy regions, the minimum flow often occurs in the winter months, whereas in other regions minimum flows occur in the season with low precipitation and large atmospheric water demand. We are not aware of any study that is explicitly analysing time series of DOYMIN.

Table 4. Continued.

Abbrev.	Commentary
DOYMAX	The timing of annual maximum streamflow can be a valuable indicator for the flood generating processes. In cold regions annual, maximum flow is often associated with snowmelt, while in other regions it may be associated with intense convective precipitation during the warm season or soil moisture. Time series of DOYMAX have for example been used to assess trends in the timing of floods in Europe (Blöschl et al., 2017) and Canada (Cunderlik and Ouarda, 2009).
DOYMIN7	Overall the interpretation of DOYMIN7 is analogous to the interpretation of DOYMIN. Note, however, that DOXMIN7 is representative of a 7-day period of sustained low flows and is less sensitive to outliers. We are not aware of any study that is explicitly analysing time series of DOYMIN.
DOYMAX7	Generally, the interpretation of DOYMAX7 is analoguous to the interpretation of DOYMAX, although DOYMAX7 represents a 1-week period of sustained high flows and is less sensitive to outliers. We are not aware of any study that is explicitly analysing time series of DOYMAX7.
GINI	The Gini coefficient is a metric that was originally established in economic sciences as a measure of economic inequality (Ceriani and Verme, 2012). It is a measure of dispersion that is not dependent on the absolute value of the variable under consideration and can be interpreted as a measure of the variability implied by the flow duration curve. It is therefore, like the CV, a relative variability measure that can easily be compared among different regions. Although we are not aware of any study investigating annual GINI time series derived from streamflow, relevant applications to observed precipitation (Rajah et al., 2014) and global hydrological model output (Masaki et al., 2014) are emerging.

3.1.3 Reliability of index values

Not all daily time steps have observations, and some daily observations have been flagged as *suspect* and were therefore removed. Consequently yearly, seasonal and monthly index values are not equally reliable. To allow users to judge the reliability of index values at individual time steps, the number of daily values used for index calculation at each time step is provided. Based on the recommendations of ECA&D13, the following rules for daily data availability can be applied to identify *reliable* index values.

1. Index values at a *yearly* time step are reliable if at least 350 daily observations are declared *reliable*.
2. Index values at a *seasonal* time step are reliable if at least 85 daily observations are declared *reliable*.
3. Index values at a *monthly* time step are reliable if at least 25 daily observations are declared *reliable*.

Note, however, that these are very conservative rules which may be relaxed depending on the needs of specific applications.

3.2 Example time series

To provide a first impression of the considered indices, Fig. 2 shows all indices at annual resolution for Wiese at Zell, located in south-western Germany. In addition, Fig. 3 shows the MEAN at monthly, seasonal and yearly resolutions of the same river.

3.3 Temporal coverage of yearly, seasonal and annual indices

Figure 4a displays the number of years covered by all considered time series, highlighting both large variations in station density and time-series length, which is consistent with the availability of the original daily time series (Do et al., 2018a). To better appraise regional differences in temporal coverage, Fig. 4b shows the distribution of the number of years that are typically available for each station for major continental regions. The median time-series length is longest for North America and Europe and shortest for Oceania and Asia. The above-mentioned daily quality control (Sect. 2) as well as ECA&D13 criteria for judging the reliability of yearly, seasonal or monthly index values (Sect. 3.1.3) imply that the space–time coverage of the index data is not equal to the coverage of the original daily time series. Figure 4c shows the distribution of the fraction of time steps that were classified as reliable for the considered continental regions and for yearly, seasonal and monthly resolutions. Overall the figure highlights that the fraction of reliable time steps is largest for the Americas, Europe and Asia, while it is lowest for Oceania and Africa. Furthermore, it should be noted that the fraction of reliable time steps is lowest for yearly indices. This is related to the fact that full years are deemed unreliable when fewer than 350 valid observations are used for computation (following the ECA&D13 rules). Note however that the relatively strict ECA&D13 rules can be relaxed and should be adapted depending on user needs.

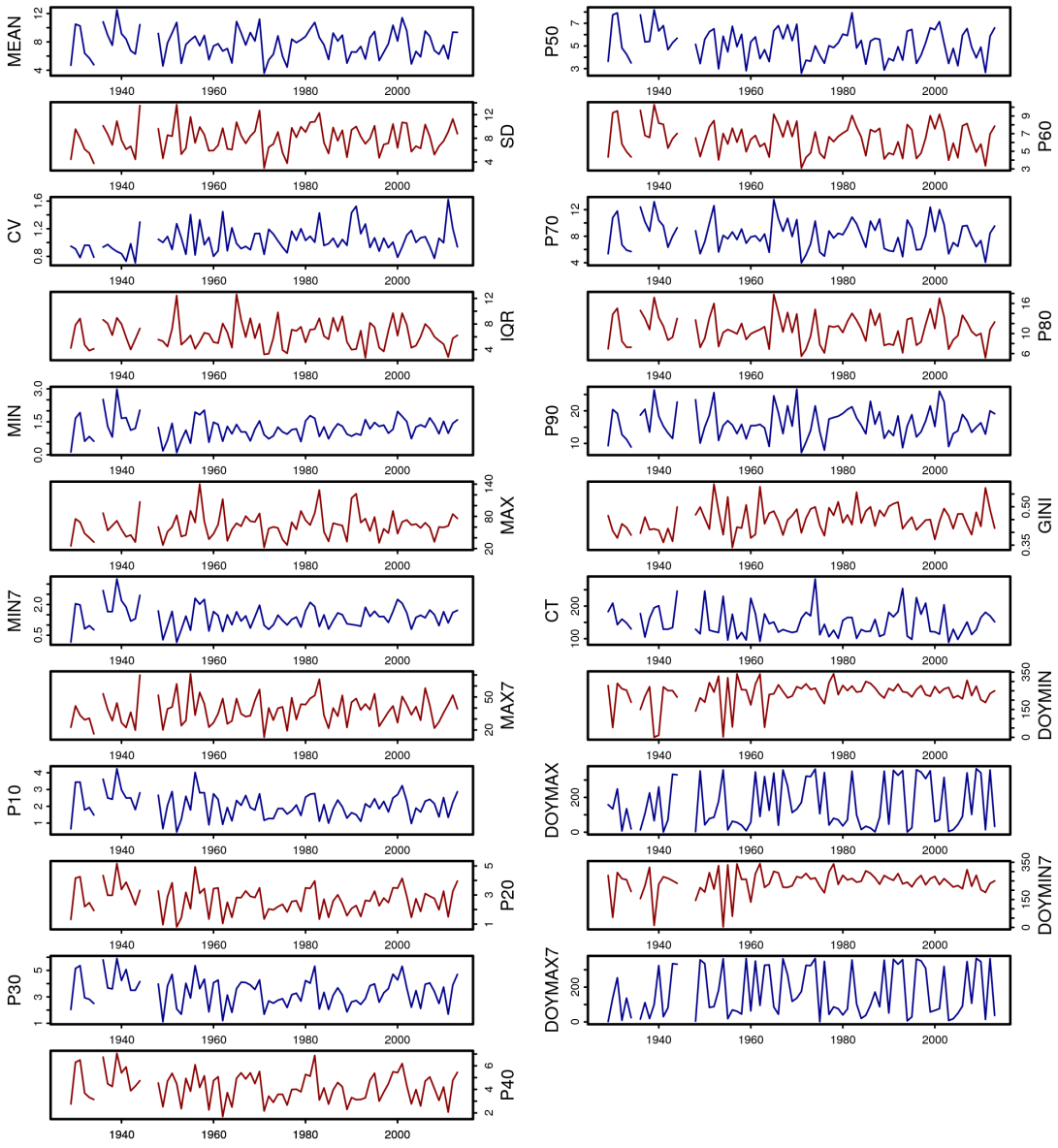


Figure 2. All considered indices at yearly resolution, shown for the River Wiese at Zell, south-western Germany. Yearly values are only displayed if they contain at least 350 reliable daily observations. See the text for details on units, interpretation and reliability classification.

4 Homogeneity assessment

4.1 Methods for homogeneity assessment

4.1.1 Homogeneity tests

Any environmental time series can be subject to inhomogeneities, i.e. unnatural sudden shifts in their statistical moments. In the simplest case, such inhomogeneities could be a jump in the mean between two time periods (see Fig. 5, top), but also changes in variability (e.g. reduced peak flows) or shifts in higher-order moments. The reasons for such inhomogeneities in streamflow time series are manifold, but they can "be related to changes in instrumentation, gauge restoration, recalibration of rating curves, flow regulation or channel engineering" (Gudmundsson and Seneviratne, 2016). As all the above-mentioned factors can be detrimental to a scientific investigation, it is essential to check time series against inhomogeneities. Here we apply a previously utilized collection of tests (Gudmundsson and Seneviratne, 2016), which is recommended by ECA&D13 and has been thoroughly tested for temperature and precipitation indices (Wijngaard et al., 2003). This collection of tests contains (i) the standard normal homogeneity test (Alexandersson, 1986), (ii) the Buishand range test (Buishand, 1982), (iii) the Pettitt test (Pettitt, 1979), and (iv) the von Neumann ratio test (von Neumann, 1941). For the application of the above-mentioned collection

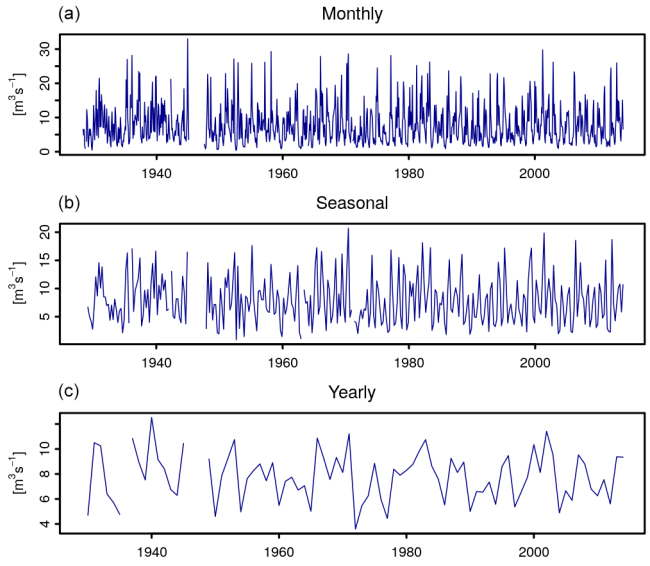


Figure 3. Monthly, seasonal, and yearly MEAN for the River Wiese at Zell, south-western Germany. Index values are only displayed if they fulfil the ECA&D13 data availability criteria. See the text for details.

of tests, we rely on tables that provide critical values of the test statistics for a given sample size that have been determined using Monte Carlo methods (ECA&D13). These tables only report critical values for a sample size of 20 and larger. Therefore, the tests can only be applied if at least 20 yearly, monthly or seasonal time steps are available. Prior to homogeneity testing, yearly, seasonal and monthly index values that are classified as unreliable according to ECA&D13 (see Sect. 3.1.3) are set to missing. Missing values were removed after pre-whitening of yearly, seasonal and monthly index time series (see Sect. 4.1.2).

4.1.2 Pre-whitening

As the considered homogeneity tests rely at least on the assumption that the data are stationary, independent and identically distributed, all indices are pre-processed (pre-whitened), aiming to reduce effects of (i) trends, (ii) seasonality, and (iii) serial correlation. For the pre-whitening procedure, linear trends and mean seasonal cycles were removed using a linear least-squares regression model which captures both the trend and the mean values as $x = b + at$, where b is the intercept, a is the trend and t is time.

1. For *yearly* indices, the linear model is fitted to and subtracted from the complete time series. This results in a time series with zero mean and no linear trend.

2. For *seasonal* indices, the linear model is fitted to and subtracted from the time series for each season (DJF, MAM, JJA, SON) individually. This results in a time series with seasonal resolution in which each season has a zero mean and no linear trend.

3. For *monthly* indices, the linear model is fitted to and extracted from the time series for each month (January, February, etc.) individually. This results in time series with monthly resolution in which each month has a zero mean and no linear trend.

As the detrended and de-seasonalized time series may still exhibit serial correlation, they were further pre-whitened by fitting a lag-1 autoregressive model and then obtaining the residuals, which are then subjected to the homogeneity analysis (Burn and Elnur, 2002; Chu et al., 2013; Gudmundsson and Seneviratne, 2016). The lag-1 autoregressive model is fitted using maximum likelihood estimation.

4.1.3 Classification of station homogeneity

To effectively combine the information of the four considered homogeneity tests, we classify the homogeneity of yearly, monthly and seasonal time-series indices following recommendations of ECA&D13:

1. *useful:* one or no tests reject the null hypothesis at the 1 % level;

2. *doubtful*: two tests reject the null hypothesis at the 1 % level;

3. *suspect*: three or four tests reject the null hypothesis at the 1 % level.

Note, however, that depending on the application, these rules may be either too relaxed or too conservative. In addition, we also introduce the following categories to account for special circumstances that can occur in this large-scale application:

4. *not sufficient data:* less than 20 yearly, seasonal or monthly reliable index values are available;

5. *constant:* all yearly, seasonal or monthly time steps have the same value;

6. *error:* an error (e.g. numerical convergence issue) occurred at any processing step.

4.2 Homogeneity testing of all yearly, seasonal and monthly time-series indices

The homogeneity analysis is applied to all indices at yearly, seasonal and monthly resolution. The rationale for applying the four tests to all indices individually is that inhomogeneities at a particular location might be relevant only for a subset of indices, while other indices are not affected. For example, it is possible that a change in instrumentation will affect peak flows, while low flows are not affected. For this homogeneity assessment, all yearly, seasonal and monthly

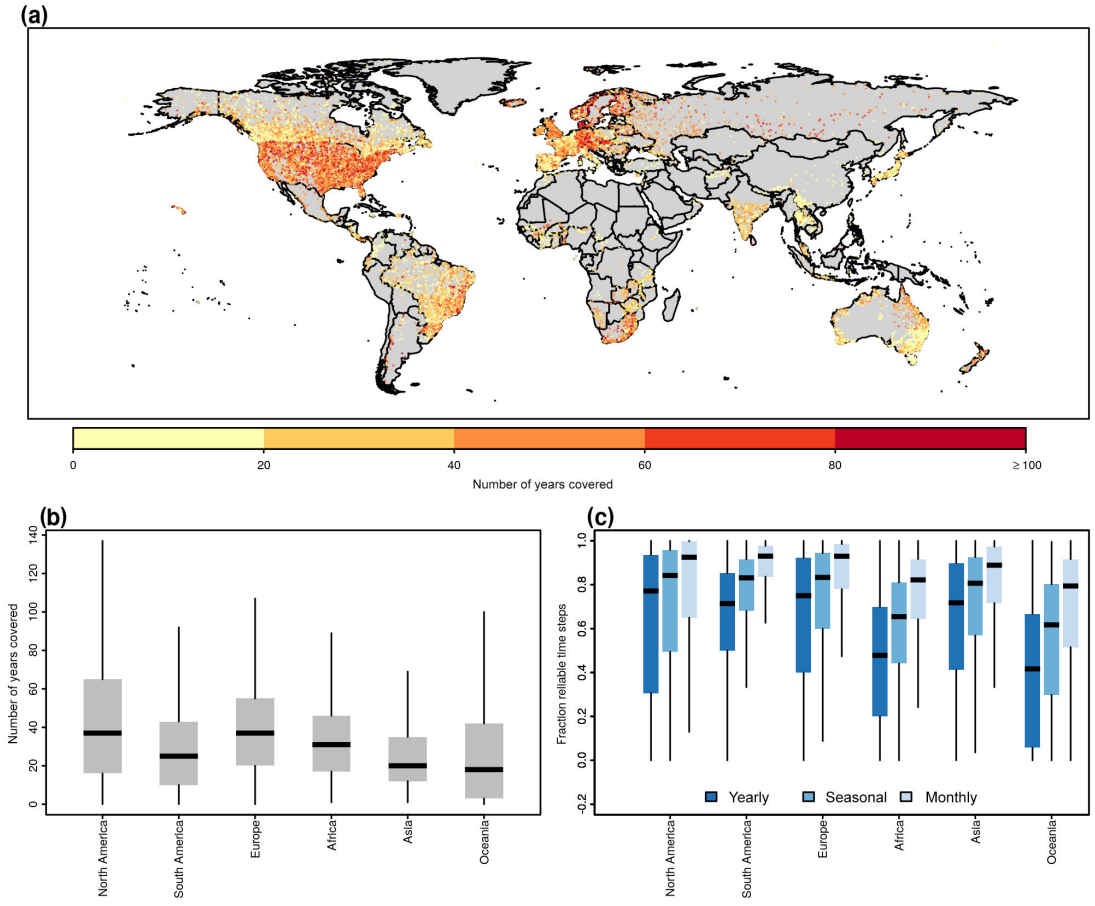


Figure 4. Temporal coverage of streamflow time-series indices. **(a)** Map of the number of years covered by each time series under consideration. **(b)** Distribution of the number of years available per time series for the continental regions of the world. **(c)** Distribution of the fraction of time steps that are classified as reliable using the ECA&D13 data availability criteria. Boxplots show the interquartile range (box) and the median (vertical bar); the whiskers extend to the most extreme point, which is not more than 1.5 times the interquartile range away from the box; outliers are omitted.

time steps that are classified as reliable (Sect. 3.1.3) are considered. This results in a conservative assessment as (i) strict data-availability criteria are applied, and (ii) because inhomogeneities could occur in a time window not relevant to a study. Therefore, the presented results can be used for a general overview of time-series homogeneity, but their suitability should always be re-considered prior to specific applications.

Figure 5 illustrates the results of the homogeneity assessment for the MEAN index for the North Umpqua River in the US. The top panel shows the monthly MEAN index, which displays a sudden jump after the first third of the record. This jump may for example be the result of upstream flow regulation and would be detrimental for climatological investigations. The lower panel shows the time series after the above-mentioned pre-whitening procedure was applied. The seasonal cycle is effectively removed and obtaining the residuals from the lag-1 autoregressive model reduced the magnitude of the sudden jump. Note also the spurious trend, which is an artefact of the de-trending that occurs in the presence of strong, sudden shifts in the mean. Nevertheless, three of the four considered tests identify this inhomogeneity at the 0.01 significance level, and the series is classified as *suspect*.

Global summaries of the number of stations in different homogeneity classes are shown in Fig. 6. Owing to the reduced number of time steps, the homogeneity testing could only be applied for approximately half of the locations at yearly resolution. Nevertheless, the homogeneity assessment highlights that the other half of the yearly indices can be considered "useful" at many locations. Only a small number of the low-flow indices (e.g. MIN, P10, P20, P30) had "constant" values and other issues were rarely detected. For both seasonal and monthly resolution, the number of stations with sufficient data for homogeneity assessment increased significantly, although it is important to recall that the homogeneity tests were in many cases applied to relatively short records (i.e. at least 20 seasons or 20 months respectively). Most of the seasonal and monthly time series with sufficient data are

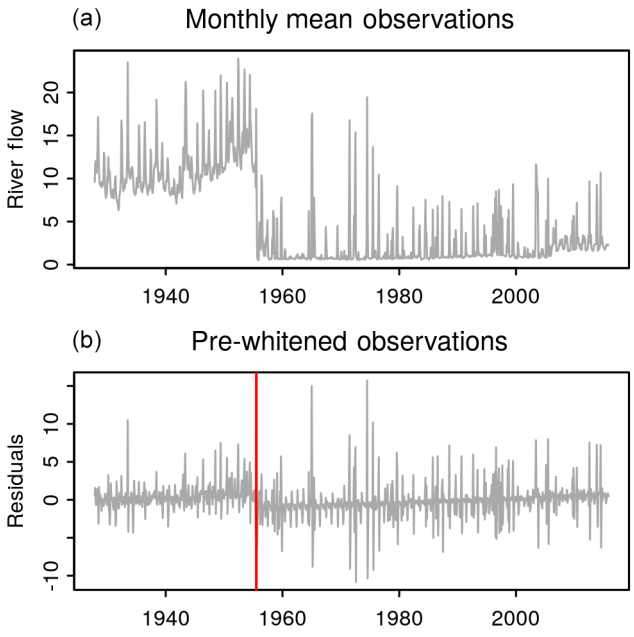


Figure 5. Homogeneity assessment of monthly mean flow of the North Umpqua River, US. **(a)** Monthly mean observations. **(b)** Pre-whitened observations together with the time step at which the standard normal homogeneity test, the Buishand range test and the Pettitt test identified a breakpoint at the 0.01 significance level.

classified as "useful", but a number of "doubtful" and "suspect" values were also detected. At a few locations, low-flow indices had constant values.

Figure 7 shows continental summaries of the homogeneity assessment at yearly, seasonal and monthly timescales and highlights the number of stations at which all indices were classified as useful according to the ECA&D13 criteria. Interestingly, the fraction of time series for which all indices have been classified as "useful" remains approximately constant irrespective of the considered time resolution. Figure 8 illustrates the effect of data availability criteria (Sect. 3.1.3) and the homogeneity assessment of the number of stations for each time step. Regardless of the temporal resolution, the number of stations reduces significantly when the homogeneity criterion is applied. This effect is more prominent at finer temporal resolution (monthly), as adding the "all indices homogenous" criterion removes approximately half of the eligible time series (bottom panel of Fig. 8). Note, however, that the presented summaries can only act as a rough guide on data availability, as criteria for including or excluding specific stations will depend on the objectives of individual future assessments.

5 Data availability and overview of the data product

5.1 Data availability

The data described in this paper are freely available as a compressed zip archive that can be downloaded from https://doi.pangaea.de/10.1594/PANGAEA.887470 (Gudmundsson et al., 2018). The zip archive contains (i) a readme file, (ii) all time-series indices and (iii) the results of all homogeneity tests. Note that the data are accompanied by additional information on the data collection process, catchment boundaries and selected catchment properties (Do et al., 2018a, b).

5.2 Time series of yearly, seasonal and annual indices

The indices derived from daily streamflow time series as described in Sects. 2 and 3 are stored in the INDICES directory. To address the different temporal resolution of the available indices (yearly, seasonal and monthly scales), the GSIM indices were organized into three respective subdirectories where each GSIM station is represented through a text file. For instance, indices at yearly resolution derived from the station with the identifier "AR_0000006" are stored as a text file called "AR_0000006.year" in the "yearly" sub-directory. Indices at seasonal and monthly resolution are stored as "AR_0000006.seas" and "AR_0000006.mon" in the respective ("seasonal", "monthly") sub-directories.

An identical data structure was adopted across all time-series files, with basic metadata (e.g. station identifier, station name, river name) stored in the header, and all index time series written in subsequent lines as a table, where (i) the first column contains the date, which is by convention the last day of the respective yearly, seasonal or monthly time step; (ii) the subsequent columns contain the index values, with column names corresponding to the abbreviations introduced in Table 4; and (iii) the last two columns contain information on the number of (missing) daily values used to compute the index.

5.3 Homogeneity of time-series indices

The results of the homogeneity analysis are stored in three tables, representing indices at yearly, seasonal and monthly resolution which are placed in the HOMOGENEITY directory and contain information on all stations. There is an identical structure for these three text files, with the first 13 columns containing important metadata such as the station identifier, name of the gauging location, and first and last time steps of the index time series. The remaining columns contain the results of four homogeneity tests that are described in the paper, and thus each index is accompanied by four columns (corresponding to the results of the (1) standard normal homogeneity test, (2) the Buishand range test, (3) the Pettitt test and (4) the Neuman ratio test).

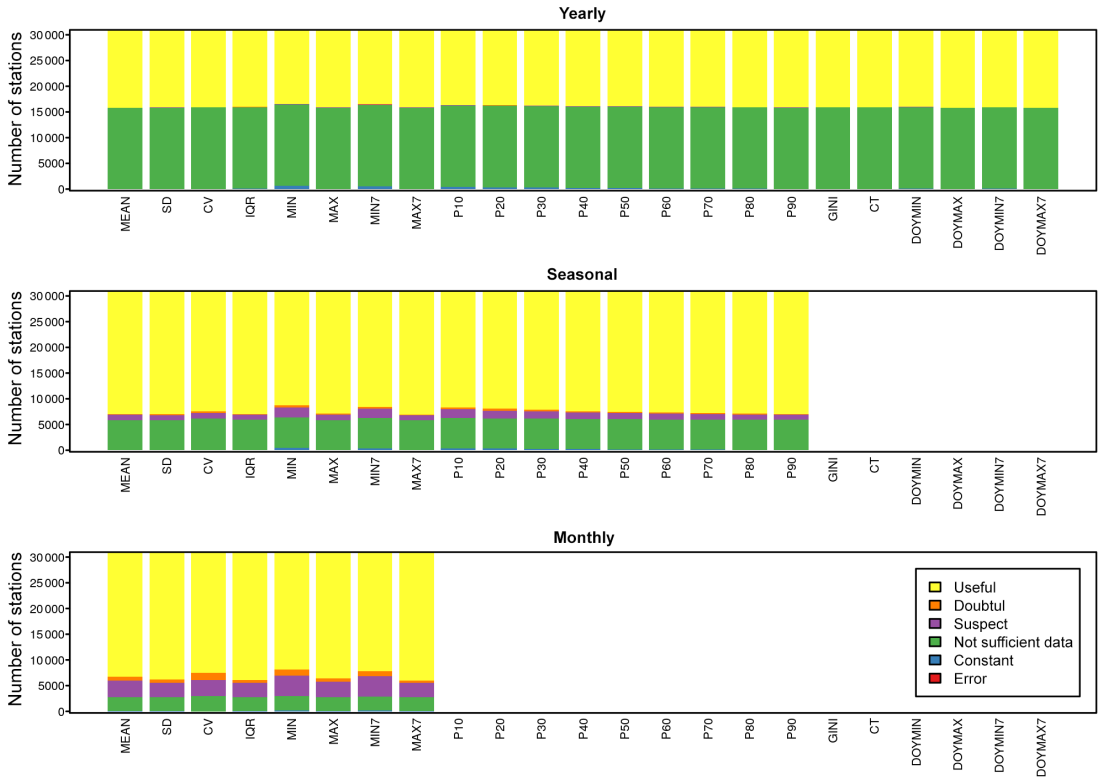


Figure 6. Global summary of the homogeneity analysis for all considered indices at yearly, seasonal and monthly resolution. Shown are the number of stations that are classified as (1) useful, (2) doubtful, (3) suspect, (4) not sufficient data, (5) constant and (6) error according to Sect. 4.1.3. Note that all six categories do occur, although some of them are rare and thus barely visible in the figure.

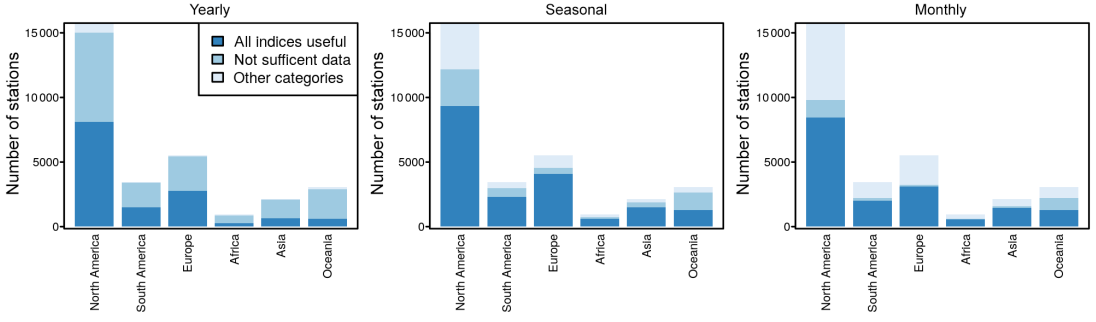


Figure 7. Continental summary of the homogeneity analysis for yearly, seasonal and monthly indices. Shown are the total number of stations at which all indices are classified as useful according to the criteria of ECA&D13, stations that did not have sufficient data for the application of the homogeneity analysis, and all other stations (other categories).

6 Summary and conclusions

Together with Do et al. (2018a) (Part 1), this paper presents the Global Streamflow Indices and Metadata Archive (GSIM), which is a unique collection of streamflow observations at more than 30 000 stations around the globe. In Part 1 (Do et al., 2018a) of the paper series we focussed on the collection and merging of freely available streamflow data worldwide. Part 1 also introduced shapefiles of catchment boundaries together with essential catchment properties such as land cover, topography and mean climatic conditions. As not all data providers allow for a free distribution of unprocessed daily values, we followed in Part 2 an approach that has been established through the ETCCDI in climate research (Klein Tank et al., 2009; Zhang et al., 2011) and introduced a set of time-series indices that can be used to assess the water balance, seasonality, low flows and floods, which are made freely available to serve the scientific community.

While focussing on time-series indices facilitates the redistribution of the data, this approach inevitably comes with

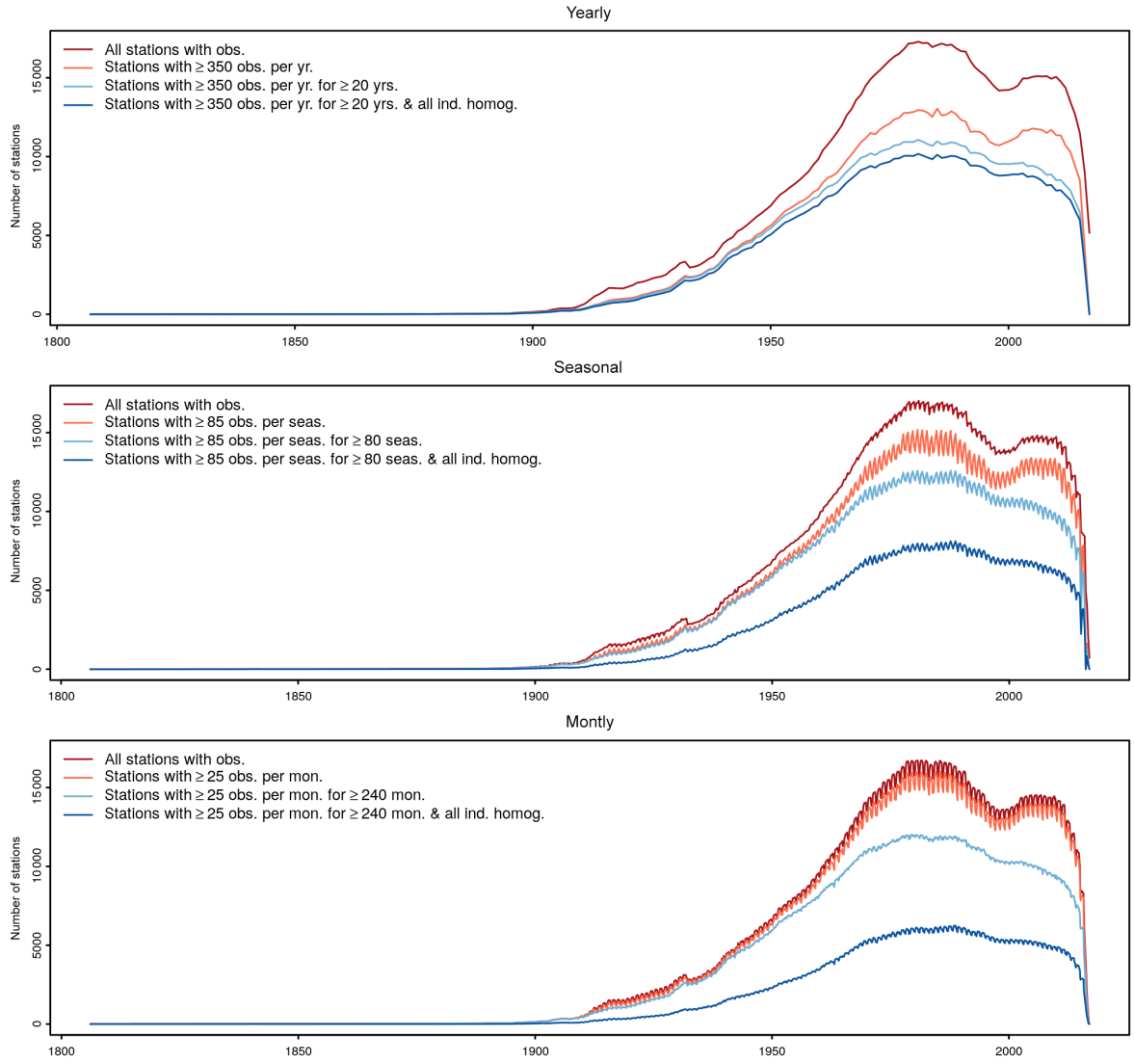


Figure 8. Temporal evolution of global station coverage, conditional on different data-selection criteria for yearly, monthly and seasonal timescales. Successively, the following criteria are applied: (i) all stations that at least one observation for the respective time step (i.e. year, season, month). (ii) Stations that have at least a critical number of observations for each time step (critical values depend on the timescale; see Sect. 3.1.3). (iii) Stations that have at least a critical number of observations for the equivalent of 20 station years (i.e. 20 yearly values, $20 \times 4 = 80$ seasonal values, $20 \times 12 = 240$ monthly values). (iv) Stations where criterion (iii) applied and all indices were considered to be useful in the homogeneity analysis (see Sect. 4.1.3).

inherent limitations. For example, many applications, including hydrological or ecological modelling, may require daily resolution data and other studies may depend on indices not included in the presented collections. Consequently, some users may prefer to seek out the original data sources (see details in Do et al., 2018a) and access the raw daily streamflow values in that manner. Nevertheless, we would like to also highlight the advantages of time-series indices: a benefit of having pre-processed the daily streamflow data into indices is that they can be readily used in studies across large regions with minimal handling of raw data files. In addition, the selected indices foster a wide variety of assessments, including water balance calculations, extreme event analysis and the identifications of trends in the world's freshwater resources.

To ensure the reliability of the published data, we first evaluated the quality of individual daily values through a combination of quality flags developed by the data providers and a transparent numerical screening approach. Subsequently, the homogeneity of yearly, seasonal and monthly indices was assessed using reproducible methods, aiming at aiding potential users to gauge the suitability of individual time series for their research questions. Note, however, that it is not the intent of this project to derive a single "best" dataset, for example, by considering a pre-defined baseline period which

gauges must cover, or by derivation of a so-called "high-quality" dataset by applying a rigorous set of quality criteria to available stations. While these approaches are of high value if a dataset is tailored to a specific application, the emphasis of GSIM is to provide a large database of streamflow observations by collating and standardizing many data sources around the world.

Given that data quality requirements can vary substantially, it will remain the work of individual users to establish selection criteria for each study, thereby finding a trade-off between data quantity (number of gauges) and data quality (record length, missing periods). While the criteria used to gauge the usability of the indices are based on the recommendations of ECA&D13, they necessarily rely on subjective decisions on what constitutes a "reliable index". For example, in some climates a gauge may be "reliable" and yet unable to provide measurements for part of the year (e.g. seasonally dry or cold climates). For this reason, attempts have been made to provide flexibility, aiming at facilitating the user to judge upon "reliability" in the context of their applications. Nonetheless, it is our hope that enabling a wide usage of streamflow indices might also lead to greater scrutiny of the data, accumulated knowledge of performance of each site and improved methods for judging the quality of streamflow observations.

There are numerous unsettled scientific questions at the global scale that this dataset has the potential to support. For example, there are unresolved questions around the relationship between trends in rainfall extremes and hydrological extremes (Do et al., 2017; Westra et al., 2013), as well as developing a better understanding of the influence of human activities on the hydrological cycle more broadly (Barnett et al., 2008; Blöschl et al., 2017; Destouni et al., 2013; Gudmundsson et al., 2017; Hegerl et al., 2015; Jaramillo and Destouni, 2015). Expanding upon recent methodological developments (Gudmundsson and Seneviratne, 2015, 2016), the newly assembled data may act as a basis for developing gridded global-scale observation-based data products. There are also likely to be many applications in fields as diverse as hydro-ecology, water quality modelling, environmental assessment and socio-hydrology. We therefore expect the presented data to be a valuable source of information to answer pending questions in global freshwater research, e.g. in the context of the World Climate Research Program Grand Challenge on Water Availability (Trenberth and Asrar, 2014) or the international research efforts on "Change in hydrology and society" (Montanari et al., 2013).

The significant increase in global gauge density and record length through the GSIM archive would not have been possible without the fact that water agencies are increasingly making data accessible online. However, the benefits of this new collection are overshadowed by challenges that are essentially bureaucratic in nature: how to systematically collate, maintain and improve streamflow data globally and who should do it. While agencies such as the GRDC would provide a natural fit for this type of task, they are currently constrained in their capacity to commit to a regular and systematic upkeep of such a global dataset. This paper series represents a one-off initiative of the authors, requiring over a year's worth of checking and evaluation and with little to no capacity for updating or extending the dataset. While it is possible that updates might be achieved through similar future efforts from the community, they are likely to be ad hoc and far from ideal. There are many troubles that can result from patchwork efforts of data collating, including (i) orphaned versions that persist in usage despite updated data being available, (ii) gauges or regions becoming out-of-sync, (iii) repeated needs to identify duplicates in overlapping datasets, (iv) information loss between versions and poor upkeep of documentation, (v) competing or "forked" databases, and many more. To remedy this situation, the hydrological community needs to collectively improve the organization of initiatives for coordinated systems that facilitate updating, storage and documentation of existing data, and to lobby for existing closed databases to be made open and accessible. As part of a global imperative for improved streamflow data, there are a number of additional activities researchers might undertake. These include (i) providing new analyses that improve the quality and understanding of the existing database; (ii) developing new automated methods that can be used systematically to maintain or improve the quality of the instrumental record; (iii) providing additional streamflow observations from missing or currently inaccessible datasets; and (iv) deriving new observational data products though better ground-truthing of remote-sensed variables, reanalysis from hydrological models or upscaling of in situ observations using machine learning.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank Sonia I. Seneviratne for the fruitful discussions on the creation of the GSIM archive. This work would not have been possible without the tremendous efforts of regional, national and international organizations in collecting and archiving river flow observations. Their work is highly appreciated.

Edited by: David Carlson
Reviewed by: Wolfgang Grabs and one anonymous referee

References

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D. B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., and

Vazquez-Aguirre, J. L.: Global observed changes in daily climate extremes of temperature and precipitation, J. Geophys. Res., 111, D05109, https://doi.org/10.1029/2005JD006290, 2006.

Alexandersson, H.: A homogeneity test applied to precipitation data, J. Climatol., 6, 661–675, 1986.

Alkama, R., Marchand, L., Ribes, A., and Decharme, B.: Detection of global runoff changes: results from observations and CMIP5 experiments, Hydrol. Earth Syst. Sci., 17, 2967–2979, https://doi.org/10.5194/hess-17-2967-2013, 2013.

Barnett, T. P., Pierce, D. W., Hidalgo, H. G., Bonfils, C., Santer, B. D., Das, T., Bala, G., Wood, A. W., Nozawa, T., Mirin, A. A., Cayan, D. R., and Dettinger, M. D.: Human-Induced Changes in the Hydrology of the Western United States, Science, 319, 1080–1083, 2008.

Beck, H. E., de Roo, A., and van Dijk, A. I. J. M.: Global maps of streamflow characteristics based on observations from several thousand catchments, J. Hydrometeorol., 16, 1478–1501, 2015.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrol. Earth Syst. Sci., 21, 2881–2903, https://doi.org/10.5194/hess-21-2881-2017, 2017.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, Earth Syst. Sci. Data, 5, 71–99, https://doi.org/10.5194/essd-5-71-2013, 2013.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., v Canjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Sraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Zivković, N.: Changing climate shifts timing of European floods, Science, 357, 588–590, 2017.

Buishand, T. A.: Some methods for testing the homogeneity of rainfall records, J. Hydrol., 58, 11–27, 1982.

Burn, D. H. and Elnur, M. A. H.: Detection of hydrologic trends and variability, J. Hydrol., 255, 107–122, 2002.

Ceriani, L. and Verme, P.: The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini, J. Econ. Inequal., 10, 421–443, 2012.

Chu, M. L., Ghulam, A., Knouft, J. H., and Pan, Z.: A Hydrologic Data Screening Procedure for Exploring Monotonic Trends and Shifts in Rainfall and Runoff Patterns, J. Am. Water Resour. As., 50, 928–942, https://doi.org/10.1111/jawr.12149, 2013.

Cunderlik, J. M. and Ouarda, T. B. M. J.: Trends in the timing and magnitude of floods in Canada, J. Hydrol., 375, 471–480, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, 2011.

Destouni, G., Jaramillo, F., and Prieto, C.: Hydroclimatic shifts driven by human water use for food and energy production, Nat. Clim. Change, 3, 213–217, 2013.

Do, H. X., Westra, S., and Leonard, M.: A global-scale investigation of trends in annual maximum streamflow, J. Hydrol., 552, 28–43, 2017.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765–785, https://doi.org/10.5194/essd-10-765-2018, 2018a.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive – Part 1: Station catalog and Catchment boundary, PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.887477, 2018b.

ECA & D Project Team and Royal Netherlands Meteorological Institute: Algorithm Theoretical Basis Document (ATBD), available at: https://www.ecad.eu/documents/atbd.pdf, 2013.

Ehsanzadeh, E. and Adamowski, K.: Trends in timing of low stream flows in Canada: impact of autocorrelation and long-term persistence, Hydrol. Process., 24, 970–980, 2010.

Fekete, B. M., Looser, U., Pietroniro, A., and Robarts, R. D.: Rationale for Monitoring Discharge on the Ground, J. Hydrometeorol., 13, 1977–1986, 2012.

Fekete, B. M., Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E., and Zhulidov, A. V.: Time for in situ renaissance, Science, 349, 685–686, 2015.

Frich, P., Alexander, L. V., Della-Marta, P., Gleason, B., Haylock, M., Tank, A. M. G. K., and Peterson, T.: Observed coherent changes in climatic extremes during the second half of the twentieth century, Clim. Res., 19, 193–212, 2002.

Gudmundsson, L. and Seneviratne, S. I.: Towards observation-based gridded runoff estimates for Europe, Hydrol. Earth Syst. Sci., 19, 2859–2879, https://doi.org/10.5194/hess-19-2859-2015, 2015.

Gudmundsson, L. and Seneviratne, S. I.: Observation-based gridded runoff estimates for Europe (E-RUN version 1.1), Earth Syst. Sci. Data, 8, 279–295, https://doi.org/10.5194/essd-8-279-2016, 2016.

Gudmundsson, L., Tallaksen, L. M., and Stahl, K.: Spatial cross-correlation patterns of European low, mean and high flows, Hydrol. Process., 25, 1034–1045, 2011.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, J. Hydrometeorol., 13, 604–620, 2012a.

Gudmundsson, L., Wagener, T., Tallaksen, L. M., and Engeland, K.: Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe, Water Resour. Res., 48, W11504, https://doi.org/10.1111/jawr.12149, 2012b.

Gudmundsson, L., Seneviratne, S. I., and Zhang, X.: Anthropogenic climate change detected in European renewable freshwater resources, Nat. Clim. Change, 7, 813–816, 2017.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Time Series Indices and Homogeneity Assessment, PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.887470, 2018.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, J. Hydrometeorol., 12, 869–884, 2011.

Hall, J., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Burlando, P., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Gaál, L., Gorbachova, L., Gül, A., Hannaford, J., Kiss, A., Kjeldsen, T., Kohnová, S., Koskela, J. J., Macdonald, N., Mavrova-Guirguinova, M., Ledvinka, O., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Osuch, M., Parajka, J., Perdigão, R. A. P., Radevski, I., Renard, B., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Blöschl, G.: A European Flood Database: facilitating comprehensive flood research beyond administrative boundaries, P. Int. Ass. Hydrol. Sci., 370, 89–95, 2015.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, Hydrol. Process., 25, 1191–1200, 2011.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, 2014.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113, D20119, https://doi.org/10.1029/2008JD010201, 2008.

Hegerl, G. C., Black, E., Allan, R. P., Ingram, W. J., Polson, D., Trenberth, K. E., Chadwick, R. S., Arkin, P. A., Sarojini, B. B., Becker, A., Dai, A., Durack, P. J., Easterling, D., Fowler, H. J., Kendon, E. J., Huffman, G. J., Liu, C., Marsh, R., New, M., Osborn, T. J., Skliris, N., Stott, P. A., Vidale, P.-L., Wijffels, S. E., Wilcox, L. J., Willett, K. M., and Zhang, X.: Challenges in quantifying changes in the global water cycle, B. Am. Meteorol. Soc., 96, 1097–1115, 2015.

Hidalgo, H. G., Das, T., Dettinger, M. D., Cayan, D. R., Pierce, D. W., Barnett, T. P., Bala, G., Mirin, A., Wood, A. W., Bonfils, C., Santer, B. D., and Nozawa, T.: Detection and Attribution of Streamflow Timing Changes to Climate Change in the Western United States, J. Climate, 22, 3838–3855, 2009.

Hisdal, H., Stahl, K., Tallaksen, L. M., and Demuth, S.: Have streamflow droughts in Europe become more severe or frequent?, Int. J. Climatol., 21, 317–333, 2001.

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., Fleig, A. K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C., and Wilson, D.: Climate-driven variability in the occurrence of major floods across North America and Europe, J. Hydrol., 552, 704–717, 2017.

Jaramillo, F. and Destouni, G.: Local flow regulation and irrigation raise global human water consumption and footprint, Science, 350, 1248–1251, 2015.

Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Resour., 25, 1287–1304, 2002.

Klein Tank, A. M. G., Zwiers, F. W., and Zhang, X.: Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation, available at: https://www.ecad.eu/documents/WCDMP_72_TD_1500_en_1.pdf, 2009.

Kumar, S., Merwade, V., Kam, J., and Thurner, K.: Streamflow trends in Indiana: Effects of long term persistence, precipitation and subsurface drains, J. Hydrol., 374, 171–183, 2009.

Kundzewicz, Z. W., Graczyk, D., Maurer, T., Piskwar, I., Radziejewski, M., Svensson, C., and Szwed, M. G.: Trend detection in river flow series: 1. Annual maximum flow, Hydrolog. Sci. J., 50, 797–810, 2005.

Lettenmaier, D. P., Wood, E. F., and Wallis, J. R.: Hydro-Climatological Trends in the Continental United States, 1948–88, J. Climate, 7, 586–607, 1994.

Lins, H. F. and Slack, J. R.: Streamflow trends in the United States, Geophys. Res. Lett., 26, 227–230, 1999.

Loon, V. and Anne, F.: Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2, 359–392, 2015.

Masaki, Y., Hanasaki, N., Takahashi, K., and Hijioka, Y.: Global-scale analysis on future changes in flow regimes using Gini and Lorenz asymmetry coefficients, Water Resour. Res., 50, 4054–4078, 2014.

McCabe, G. J. and Wolock, D. M.: A step increase in streamflow in the conterminous United States, Geophys. Res. Lett., 29, 2185–2185, 2002.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, 8th Conference on Applied Climatology, 179–184, 1993.

Milly, P. C. D., Dunne, K. A., and Vecchia, A. V.: Global pattern of trends in streamflow and water availability in a changing climate, Nature, 438, 347–350, 2005.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., and Belyaev, V.: "Panta Rhei-Everything Flows": Change in hydrology and society – The IAHS Scientific Decade 2013–2022, Hydrolog. Sci. J., 58, 1256–1275, 2013.

Moore, J. N., Harper, J. T., and Greenwood, M. C.: Significance of trends toward earlier snowmelt runoff, Columbia and Missouri Basin headwaters, western United States, Geophys. Res. Lett., 34, L16402, https://doi.org/10.1029/2007GL031022, 2007.

Oki, T. and Kanae, S.: Global Hydrological Cycles and World Water Resources, Science, 313, 1068–1072, 2006.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Res. Appl., 19, 101–121, 2003.

Oliveira, P. J. C., Davin, E. L., Levis, S., and Seneviratne, S. I.: Vegetation-mediated impacts of trends in global radiation on land hydrology: a global sensitivity study, Glob. Change Biol., 17, 3453–3467, 2011.

Pettitt, A. N.: A Non-Parametric Approach to the Change-Point Problem, J. Roy. Stat. Soc. C-App., 28, 126–135, 1979.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M.: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century, J. Climate, 29, 4083–4097, 2016.

Rajah, K., OĹeary, T., Turner, A., Petrakis, G., Leonard, M., and Westra, S.: Changes to the temporal distribution of daily precipitation, Geophys. Res. Lett., 41, 8887–8894, 2014.

Rauscher, S. A., Pal, J. S., Diffenbaugh, N. S., and Benedetti, M. M.: Future changes in snowmelt-driven runoff timing over the western US, Geophys. Res. Lett., 35, L16703, https://doi.org/10.1029/2008GL034424, 2008.

Reek, T., Doty, S. R., and Owen, T. W.: A Deterministic Approach to the Validation of Historical Daily Temperature and Precipitation Data from the Cooperative Network, B. Am. Meteorol. Soc., 73, 753–762, 1992.

Regonda, S. K., Rajagopalan, B., Clark, M., and Pitlick, J.: Seasonal Cycle Shifts in Hydroclimatology over the Western United States, J. Climate, 18, 372–384, 2005.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., and Carrillo, G.: Characterizing hydrologic change through catchment classification, Hydrol. Earth Syst. Sci., 18, 273–285, https://doi.org/10.5194/hess-18-273-2014, 2014.

Shiklomanov, I. A., Babkin, V. I., Penkova, N. V., Georgievsry, V. Y., Zaretskaya, I. P., Izmailova, A. V., Balonishnikova, J. A., Grigorkina, T. E., Grube, T. V., Skoryatina, K. V., Tsytsenko, and Yunitsyna, V. P.: World Water Resources at the Beginning of the Twenty-First Century, Cambridge University Press, 2004.

Shukla, S. and Wood, A. W.: Use of a standardized runoff index for characterizing hydrologic drought, Geophys. Res. Lett., 35, L02405, https://doi.org/10.1029/2007GL032487, 2008.

Sippel, S., Zscheischler, J., Heimann, M., Otto, F. E. L., Peters, J., and Mahecha, M. D.: Quantifying changes in climate variability and extremes: Pitfalls and their overcoming, Geophys. Res. Lety., 42, 9990–9998, 2015.

Small, D., Islam, S., and Vogel, R. M.: Trends in precipitation and streamflow in the eastern US: Paradox or perception?, Geophys. Res. Lett., 33, L03403, https://doi.org/10.1029/2005GL024995, 2006.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J.: Streamflow trends in Europe: evidence from a dataset of near-natural catchments, Hydrol. Earth Syst. Sci., 14, 2367–2382, https://doi.org/10.5194/hess-14-2367-2010, 2010.

Stahl, K., Tallaksen, L. M., Hannaford, J., and van Lanen, H. A. J.: Filling the white space on maps of European runoff trends: estimates from a multi-model ensemble, Hydrol. Earth Syst. Sci., 16, 2035–2047, https://doi.org/10.5194/hess-16-2035-2012, 2012.

Stewart, I. T., Cayan, D. R., and Dettinger, M. D.: Changes toward Earlier Streamflow Timing across Western North America, J. Climate, 18, 1136–1155, 2005.

Svensson, C., Kundzewicz, W. Z., and Maurer, T.: Trend detection in river flow series: 2. Flood and low-flow index series, Hydrolog. Sci. J., 50, 811–824, 2005.

Tallaksen, L. M. and van Lanen, H. A. J.: Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater, Elsevier, 2004.

Tallaksen, L. M., Madsen, H., and Clausen, B.: On the definition and modelling of streamflow drought duration and deficit volume, Hydrolog. Sci. J., 42, 15–33, 1997.

Trenberth, K. E. and Asrar, G. R.: Challenges and Opportunities in Water Cycle Research: WCRP Contributions, Surv. Geophys., 35, 515–532, 2014.

Vogel, R. M. and Fennessey, N. M.: Flow-Duration Curves. I: New Interpretation and Confidence Intervals, J. Water Res. Plan. Man., 120, 485–504, 1994.

von Neumann, J.: Distribution of the Ratio of the Mean Square Successive Difference to the Variance, Ann. Math. Stat., 12, 367–395, 1941.

Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B.: Global Water Resources: Vulnerability from Climate Change and Population Growth, Science, 289, 284–288, 2000.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resour. Res., 52, 1847–1865, 2016.

Westra, S., Alexander, L. V., and Zwiers, F. W.: Global Increasing Trends in Annual Maximum Daily Precipitation, J. Climate, 26, 3904–3918, 2013.

Wijngaard, J. B., Klein Tank, A. M. G., and Können, G. P.: Homogeneity of 20th century European daily temperature and precipitation series, Int. J. Climatol., 23, 679–692, 2003.

Zaitchik, B. F., Rodell, M., and Olivera, F.: Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme, Water Resour. Res., 46, W06507, https://doi.org/10.1029/2009WR007811, 2010.

Zhang, X., Harvey, K. D., Hogg, W. D., and Yuzyk, T. R.: Trends in Canadian streamflow, Water Resour. Res., 37, 987–998, 2001.

Zhang, X., Hegerl, G., Zwiers, F. W., and Kenyon, J.: Avoiding Inhomogeneity in Percentile-Based Indices of Temperature Extremes, J. Climate, 18, 1641–1651, 2005.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data, Wires. Clim. Change, 2, 851–870, 2011.

Supplementary Article 2

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., & Seneviratne, S. I. (2019). Observed trends in global indicators of mean and extreme streamflow. Geophysical Research Letters, 46. https://doi.org/10.1029/2018GL079725.

Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL079725

Observed Trends in Global Indicators of Mean and Extreme Streamflow

L. Gudmundsson[1], M. Leonard[2], H. X. Do[2,3], S. Westra[2], and S. I. Seneviratne[1]

[1]Institute for Atmospheric and Climate Science, Department of Environmental Systems Science, ETH, Zurich, Switzerland, [2]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia, Australia, [3]Faculty of Environment and Natural Resources, Nong Lam University, Ho Chi Minh City, Vietnam

Key Points:
- A global assessment of trends in streamflow covering 1951–2010 and 14 subcontinental regions
- The significance of regional trends is assessed, revealing complex spatiotemporal change patterns
- Indicators of mean and extreme (low and high) streamflow often share the same sign of change

Supporting Information:
- Supporting Information S1

Correspondence to:
L. Gudmundsson,
lukas.gudmundsson@env.ethz.ch

Citation:
Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., & Seneviratne, S. I. (2019). Observed trends in global indicators of mean and extreme streamflow. *Geophysical Research Letters*, *46*. https://doi.org/10.1029/2018GL079725

Received 22 JUL 2018
Accepted 20 DEC 2018
Accepted article online 26 DEC 2018

Abstract This study investigates global changes in indicators of mean and extreme streamflow. The assessment is based on the Global Streamflow Indices and Metadata archive and focuses on time series of the annual minimum, the 10th, 50th, and 90th percentiles, the annual mean, and the annual maximum of daily streamflow. Trends are estimated using the Sen-Theil slope, and the significance of mean regional trends is established through bootstrapping. Changes in the indices are often regionally consistent, showing that the entire flow distribution is moving either upward or downward. In addition, the analysis confirms the complex nature of hydrological change where drying in some regions (e.g., in the Mediterranean) is contrasted by wetting in other regions (e.g., North Asia). Observed changes are discussed in the context of previous results and with respect to model estimates of the impacts of anthropogenic climate change and human water management.

Plain Language Summary Studies of trends in streamflow data from across the globe are essential for understanding patterns and changes in water availability (e.g., regions of deficit and abundance) and evaluating the fidelity of global water availability models. This study evaluates historical trends in streamflow data, using a new data set of observations from over 30,000 sites around the world. The study is comprehensive, looking at changes in low flows (defined as the lowest day of flow in each year), average flows, and high flows (the highest day of flow in each year). An interesting outcome is that where trends are present in a region, the direction of the trend is often consistent across all indices for that region (consistently drier or wetter), as distinct from the possibility of stronger extremes (wetter maximums and drier minimums).

1. Introduction

Among the most important implications of anthropogenic climate change are the potential for both large-scale changes in water availability (Greve et al., 2018; Schewe et al., 2014) and increases in the magnitude and occurrence of floods and droughts (Hirabayashi et al., 2013; Prudhomme et al., 2014). Simultaneously, the unprecedented scale of on-ground human interventions in the water cycle—including reservoir construction, irrigation, and land cover change—is also affecting terrestrial hydrology and might even exceed the impact of future climate change in some regions (Haddeland et al., 2014).

To better anticipate future changes in the world's water resources and hydrological extremes, it is essential to analyze already observed changes. Among all components of the terrestrial water cycle, streamflow (including river flow) is arguably the variable that has been monitored with the highest station density and the longest temporal coverage (Fekete et al., 2012, 2015; Hannah et al., 2011) and is thus the best we have for investigating past changes in water resources and hydrological extremes.

An increasing number of regional studies have drawn a complex picture of trends in annual streamflow statistics over several (sub)continents, including North America (Burn & Elnur, 2002; Douglas et al., 2000; Hodgkins et al., 2017; Lettenmaier et al., 1994; Lins & Slack, 1999; Mallakpour & Villarini, 2015; McCabe & Wolock, 2002; McClelland et al., 2006), South America (Genta et al., 1998; Marengo et al., 1998; Pasquini & Depetris, 2007), Europe (Blöschl et al., 2017; Gudmundsson et al., 2017; Hannaford et al., 2013; Hisdal et al., 2001; Hodgkins et al., 2017; Stahl et al., 2010), and Asia (Adam & Lettenmaier, 2008; MacDonald et al., 2007; McClelland et al., 2006; Tananaev et al., 2016). It is, however, difficult to generalize from these assessments, as they are often tailored to match conditions in a specific continent, consider

different periods, and have variations in method and selected indices. Furthermore, little work has been published for several important landmasses, including Africa and large parts of Eurasia.

Of the global studies that focus on trends in observed streamflow, some are dedicated to changes in the total freshwater fluxes to the ocean, thereby focusing on the outlets of continental-scale river basins (Alkama et al., 2011; Dai et al., 2009; Dai & Trenberth, 2002; Labat et al., 2004; Milliman et al., 2008). Consistent with the regional studies, these highlight spatially complex trend patterns. Although these assessments are of high relevance for ocean and Earth system dynamics, they focus on the net terrestrial water balance and cannot infer regional- to local-scale changes.

Another branch of global studies has assessed streamflow trends of individual water bodies. Some studies have focused on investigating changes in a few carefully selected large river basins (Jaramillo & Destouni, 2015; Kundzewicz et al., 2005; Milly et al., 2005; Svensson et al., 2005), thereby taking advantage of better quality control of the individual records but suffering from relatively small sample sizes and sparse spatial coverage. This is contrasted by other investigations that take advantage of large samples of available time series with sufficient observations (Berghuijs et al., 2017; Do et al., 2017), thereby providing a richer spatial picture of changes in water availability.

As for regional studies, there is a large heterogeneity between the individual global assessments, including a wide range of research questions, different spatial sampling schemes, and different time periods. Some studies are dedicated to investigating mean flows (Alkama et al., 2011; Dai et al., 2009; Dai & Trenberth, 2002; Jaramillo & Destouni, 2015; Labat et al., 2004; Milliman et al., 2008; Milly et al., 2005), while others focus on floods (Berghuijs et al., 2017; Do et al., 2017; Kundzewicz et al., 2005; Svensson et al., 2005) or low-flow indicators (Svensson et al., 2005). In summary, the heterogeneity of past global-scale assessments makes it difficult to draw generalized conclusions on observable changes of streamflow around the world.

This study updates previous assessments of worldwide changes in streamflow, using a database with unprecedented spatial coverage of streamflow observations and using indicators of low, mean, and high flows. To account for regional differences in data availability, trends are analyzed for three overlapping 40-year periods from 1951 to 2010, maximizing the spatiotemporal coverage of the investigation. Finally, the significance of the observed trends is established at the subcontinental scale.

2. Data

Streamflow observations are taken from the Global Streamflow Indices and Meta data archive (GSIM; Do et al., 2018b; Gudmundsson et al., 2018b), which is available in the public domain (Do et al., 2018a; Gudmundsson et al., 2018a) and holds information from more than 30,000 gauging stations. Annual time series information is available through indices computed from daily values that represent a wide range of flow properties at monthly, seasonal, and yearly resolution. Here the following indices are considered:

1. Low flows are represented through time series of the annual minimum (MIN) and the annual 10th percentile (P10).
2. Average flow conditions are characterized using the annual mean (MEAN) and the annual 50th percentile/median (P50).
3. High flows are represented through time series of the annual maximum (MAX) and the annual 90th percentile (P90).

Daily time series used to compute the GSIM indices underwent a formal quality assessment (Gudmundsson et al., 2018b). The assessment utilized quality flags from individual data providers and automated screening methods that flag implausible values. Only daily records that passed this assessment were used for index calculation. Because the extremal indices (MIN, P10, P90, and MAX) are sensitive to data availability, years with less than 350 valid daily observations were set to missing for each station, as recommended by ECA, & D Project Team, and K. Royal Netherlands Meteorological Institute (2013; hereafter ECA&D13).

Note also that GSIM combines information from all gauging stations from the contributing data bases. Consequently, both near-natural and regulated catchments are included (Do et al., 2018b; Gudmundsson et al., 2018b). In this study no attempt is made to distinguish between these cases. Instead, trends in the complete observational record are documented, as changes in atmospheric boundary conditions and human water management might both trigger changes in streamflow.

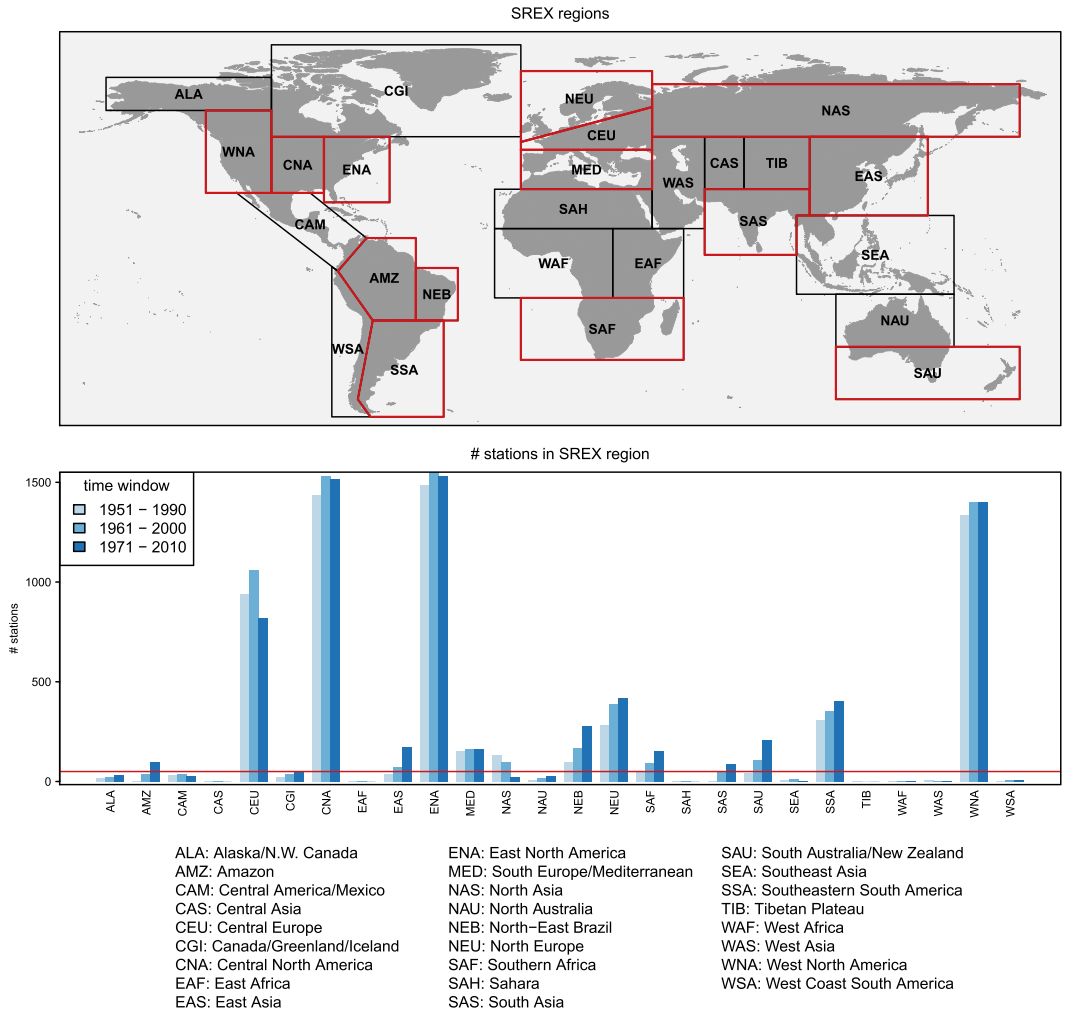


Figure 1. Subcontinental regions defined by the Special Report On Extremes (SREX; Seneviratne et al., 2012). Top: world map of all regions, where regions with more than 50 stations with sufficient data in at least one of the three considered 40-year periods are highlighted in red. Bottom: number of stations with sufficient data for each period and each SREX region. The red line indicates the 50-station threshold.

Given the complex nature of in situ observations entering the GSIM archive, spatial and temporal coverage of the considered streamflow time series varies substantially around the globe. Therefore, and because trends can be influenced by decadal variability, the following 40-year periods were analyzed: 1951–1990, 1961–2000, and 1971–2010. Based on previously suggested data availability criteria (ECA&D13) for trend analysis, only stations where at least 70% of the years are available were considered. This criterion was applied to each of the 40-year periods separately. As a result, the spatial coverage differs across the periods.

The significance of trends is evaluated at the subcontinental scale by grouping stations into 26 regions that were designed for analyzing regional climate change and are defined in the Special Report on Extremes of the Intergovernmental Panel on Climate Change (Seneviratne et al., 2012; later referred to as SREX regions). Figure 1 shows the SREX regions alongside the number of stations that fulfill the data availability criteria for each region and each 40-year period. Only regions and periods with at least 50 stations were considered for subcontinental-scale assessment.

3. Trend Estimation

Following previous studies (Stahl et al., 2010, 2012), trends at individual stations were computed using the robust Sen-Theil slope estimator (Sen, 1968). To make trend estimates from catchments with different sizes

and from different climates comparable, they are expressed in units of percent change per decade (i.e., 10 years; Stahl et al., 2012), such that

$$T_s = \frac{\tau_s \times 10 \text{ years}}{\bar{x}_s} \times 100 \quad (1)$$

where T_s is the trend at location s in units of percentage change per decade, τ_s is the Sen-Theil slope estimator, and $\bar{x}_s$ is the mean of the index time series.

To be able to detect changes at the level of the SREX regions, a resampling method is proposed that accounts for within-region spatial dependence (Burn & Elnur, 2002; Douglas et al., 2000; Wilks, 2011).

The regional trend test is as follows:

1. For a given region, compute the regional trend, $\overline{T_s}$ defined as the average of all T_s in that region.
2. Repeat 2,000 times:

2.1 Resample with replacement the year order of all data within the region while maintaining the spatial dependence within individual years, following the procedure described in Burn and Elnur (2002).
2.2 Compute at each location T_s^*, the trend expressed in percent change per decade of the resampled time series.
2.3 Compute the regional trend, $\overline{T}_s^*$, as the regional mean of T_s^*.

3. Estimate p, that is, the probability of $\overline{T_s}$ on the distribution of $\overline{T}_s^*$ (the bootstrap distribution) as the fraction of $\overline{T_s} < \overline{T}_s^*$.

Significance of regional trends is reported at the $p < 0.01$ and $p < 0.1$ level for negative trends and at the $p > 0.9$ and $p > 0.99$ level for positive trends.

Note that this procedure is related to previously suggested closed-form (Helsel & Frans, 2006) and resampling-based (Douglas et al., 2000) regional adaptations of the Mann-Kendal trend test. However, the method introduced here does not require the additional step of computing the Mann-Kendall statistic. Instead, it operates on the variable of interest, the regional trend ($\overline{T_s}$). We note that regional testing procedures have the inherent limitation that they cannot consider subregional variability, with the potential for groups of stations with positive and negative trends to mask each other out.

4. Results and Discussion

4.1. Overview

Figure 2 maps the trend magnitude of the time series indices for each period. Visual inspection of the results highlights that streamflow is not changing uniformly around the world and that the considered period can have significant effects on both the sign and magnitude of the trend.

To better understand the nature of the observed trends, Figure 3 shows the regional trends, which often point in the same direction across all indices. In the following, these regional changes will be summarized and discussed in the context of selected observational studies. In addition, the observed change patterns will be put into the context of model projections of water availability (precipitation minus evapotranspiration; Greve et al., 2018) and runoff (Haddeland et al., 2014). Note that the aforementioned studies are based on different model ensembles, each having their distinct characteristics. Greve et al. (2018) is based on the CIMIP5 ensemble (Taylor et al., 2011) that includes a large sample of global climate models but does not account for human water management and land cover change. Conversely, Haddeland et al. (2014) is based on the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) Fast Track ensemble (Warszawski et al., 2014) of global hydrological models driven with selected global climate models that accounts for both climate change and on-ground human activities. Finally, it is noted that ocean-atmosphere oscillations can be an important influence on decadal streamflow variability that have been studied elsewhere in great detail, including, for example, global (Wanders & Wada, 2015; Ward et al., 2010), North America (Burn, 2008; Tootle et al., 2005; Tootle & Piechota, 2006), Europe (Bouwer et al., 2006, 2008; Kingston et al., 2012), and Australia (Kuhnel et al., 1990; Verdon et al., 2004) assessments.

4.2. North America

North America has the highest number of stations of all the continents considered. In West North America (WNA) there is no consistent change pattern. Increasing regional average low flows (MIN

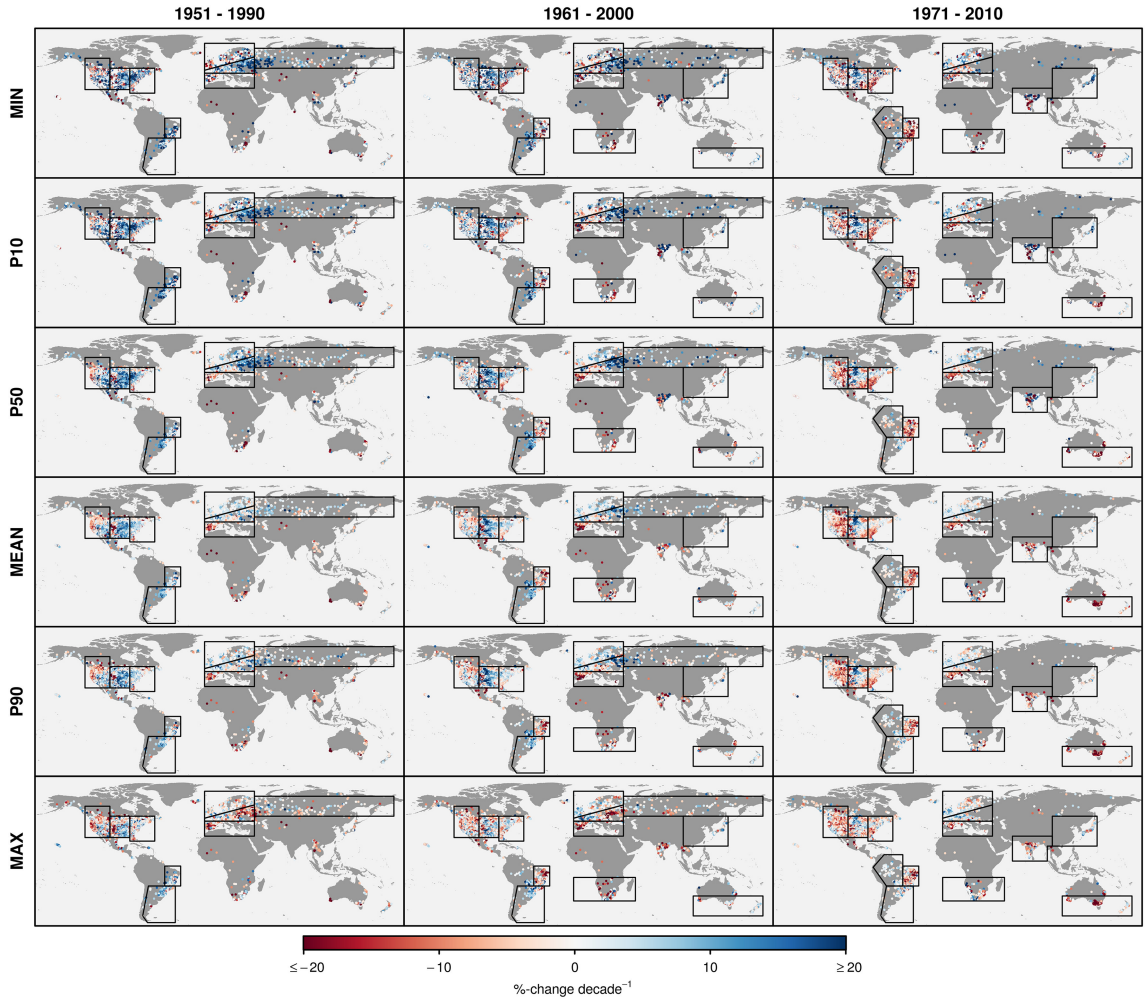


Figure 2. Trends in annual indicators of mean and extreme streamflow. Columns represent three 40-year periods. Rows represent the annual minimum (MIN), the annual 10th percentile (P10), the annual 50th percentile (P50), the annual MEAN (MEAN), the annual 90th percentile (P90), and the annual maximum (MAX). SREX regions with at least 50 stations with sufficient data are highlighted. See supporting information for high-resolution maps.

and P10) are detected for the 1951–1990 and 1961–2000 periods, whereas decreasing mean annual flows (MEAN) and high flows (P90) are found in the 1971–2010 period. In contrast, streamflow has increased significantly in Central North America (CNA) throughout the first two periods (1951–1990 and 1961–2000) across all indices, except for the annual maximum flow (MAX). This wetting tendency weakens in 1971–2010, where only MIN and P10 show a significantly increasing regional trend. There has been an increase in streamflow in East North America (ENA) over the first two periods (1951–1990 and 1961–2000), which is most pronounced for low flows and mean flows but less pronounced for high flows. In 1971–2010, the change pattern reversed, with a significantly declining regional trend for all indices except P90.

Overall, the results confirm previous assessments that focus on observations prior to the year 2000 in the United States (Douglas et al., 2000; Lettenmaier et al., 1994; Lins & Slack, 1999; McCabe & Wolock, 2002). These studies emphasize the tendency for increasing low and mean flows throughout the region, which is particularly pronounced in the central north of the U.S. There is also agreement in the lack of observed annual maximum trends, although only a few studies have focused on the period after 2000 (Hodgkins et al., 2017; Mallakpour & Villarini, 2015).

Based on simulations of the CMIP5 ensemble, Greve et al. (2018) report a clear tendency toward wetter conditions in WNA and ENA with no clear change pattern in CNA. Conversely, Haddeland et al. (2014) report a

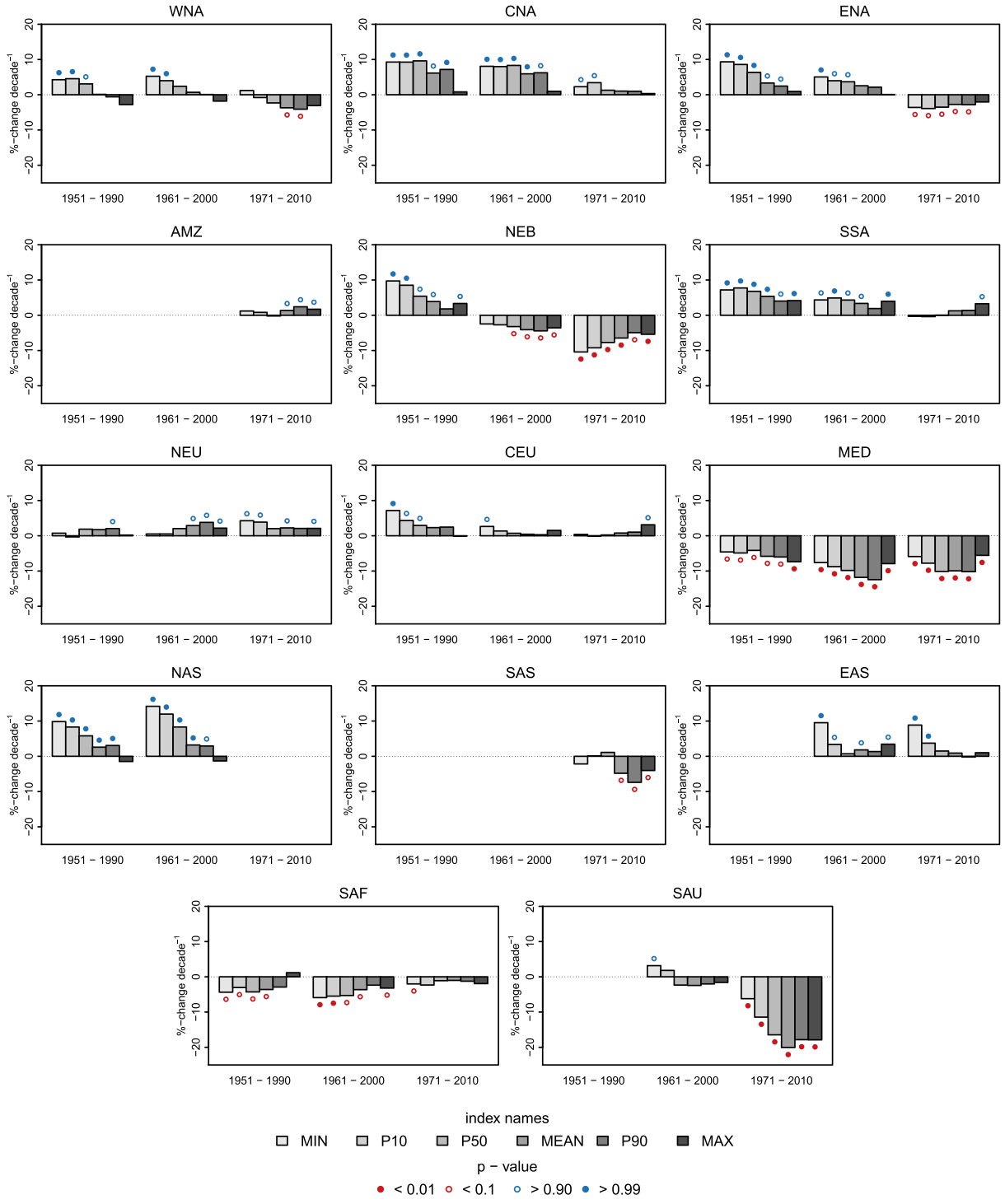


Figure 3. Regional trends for SREX regions with at least 50 stations in one of the considered periods. Regional trends are computed for all indicators of mean and extreme streamflow. Significance of the regional trend is reported as the probability of the observed trend on the bootstrap distribution. Regional trends are only provided for periods with at least 50 stations.

tendency for decreasing water availability, especially in the south of North America, which is triggered by human water and land management as simulated in the considered ISIMIP model ensemble. None of these are directly comparable with the observed change patterns, which exhibit shifting signs in both WNA and ENA throughout the study period.

4.3. South America

In South America, data availability increases throughout the study period. In the Amazon region (AMZ) only the 1971–2010 period has more than 50 stations available. In this period, the median (P50) and high flows (P90 and MAX) show a significant increasing regional trend. In North-East Brazil (NEB) all indices except P10 show a significant increasing regional trend in the 1951–1990 period. This pattern reverses thereafter, and all indices exhibit a negative regional trend for 1961–2000, although not all are significant. In the 1971–2010 period all indices show a significant declining regional trend. Southeastern South America (SSA) has wetting trends in all indices in the first period. In the second period all indices except P90 also show significant increasing trends, but this increasing tendency comes to an abrupt stop in the latter period, where no significant trends are found in most of the indices.

Analyzing streamflow observations in the second half of the twentieth century, Marengo et al. (1998) did find mostly positive trends in the region comprising parts of AMZ and NEB. This is partly consistent with the presented results but does not report the reversing trend pattern in NEB occurring between the first and second periods. For rivers draining to the south Atlantic, Pasquini and Depetris (2007) report complex spatiotemporal trend patterns using observations up to the early 2000s. For a region similar to SSA, Genta et al. (1998) report increasing discharge trends for the second half of the twentieth century with a tendency to level off, which is in agreement with the presented results.

Overall, climate models from the CMIP5 ensemble suggest that both AMZ and NEB have a tendency for becoming drier with increasing global mean temperatures, while SSA is likely to become wetter (Greve et al., 2018). However, an alternative ensemble suggests that SSA might also become increasingly drier in a warmer climate (Haddeland et al., 2014). On-ground water and land management is estimated to only have a limited impact on freshwater resources in most parts of South America (Haddeland et al., 2014).

4.4. Europe

Europe is among the best monitored regions with respect to streamflow around the world. In North Europe (NEU), only P90 exhibited a weak increasing regional trend in the 1951–1990 period. In 1961–2000 all indices show a weak incline, while only MEAN, P90, and MAX are significant. In the 1971–2010 period all indices showed a weak increasing trend, with significant MIN, P10, MEAN, and MAX suggesting a slight upward shift of the annual daily streamflow distribution. In Central Europe (CEU), there is a significant upward regional trend in MIN, P10, and P50 in the first period (1951–1990). In the subsequent period (1961–2000) almost no changes occur except for MIN, which shows a weak positive regional trend. In the last period (1971–2010) only MAX shows a weak significant regional trend in CEU. The South Europe/Mediterranean (MED) region (note that all considered stations are in Europe) shows the strongest and the most consistent pattern of the entire study. Here all indices show strong and significant declining regional trends throughout all considered time periods, highlighting an overall reduction of freshwater availability in this region.

Studies focusing on trends in flood frequency (Hodgkins et al., 2017) and drought indicators (Hisdal et al., 2001) report that there is little evidence for changes in these quantities in Europe. However, several studies have documented the tendency for drying in the South of Europe and increasingly wet conditions in the north (Stahl et al., 2010, 2012). Through linking this observational pattern to historical climate model simulations, the observed trends in pan-European freshwater availability have been attributed to anthropogenic climate change (Gudmundsson et al., 2017).

Future climate projections indicate a continuation of the observed trend pattern in Europe with increasingly dry conditions in MED, wetting conditions in NEU, and little change in CEU (Greve et al., 2018; Haddeland et al., 2014). In addition, Haddeland et al. (2014) indicate that human water and land management may also contribute to declining streamflow values in southern Europe, which might have amplified the observed strong negative trend throughout all aspects of the flow distribution.

4.5. Asia

Spatiotemporal data availability is variable over the Asian continent. In North Asia (NAS), all indices except MAX show a significant increasing regional trend for the 1951–1990 and 1961–2000 periods. In the last period (1971–2010) there are less than 50 stations in this large region. In South Asia (SAS) only the 1971–2010

period passes the data availability criteria, having a significantly declining regional trend in MEAN, P90, and MAX. Relatively better data coverage is found in East Asia (EAS), with sufficient data to cover the last two time periods (note that most stations are in Japan). For 1961–2000, MIN, P10, MEAN, and MAX show significant increasing regional trends. For 1971–2010, the regional trends of MIN and P10 are increasing significantly.

Several previous studies have documented increasing streamflow trends in the north of the Asian continent (Adam & Lettenmaier, 2008; MacDonald et al., 2007; McClelland et al., 2006; Tananaev et al., 2016), which appears to persist past the year 2000 and is also visible in hydrological extremes (Tananaev et al., 2016). A regional study in the Indian subcontinent confirms the tendency toward decreased water availability and attributes it to anthropogenic climate change (Mondal & Mujumdar, 2012).

Global climate models project that water availability will increase in NAS as a consequence of global warming (Greve et al., 2018). This is consistent with the observational results of the present study, and only limited impacts of human management on water resources is expected (Haddeland et al., 2014). In SAS climate models indicate that global warming will increase water availability (Greve et al., 2018), contrasting our observational findings. However, the simulations assessed by Haddeland et al. (2014) suggest that human land management is reducing runoff in the Indian subcontinent offering a possible explanation for the observed signal. Also in EAS, climate models project increasing water availability in an warming climate (Greve et al., 2018) and impacts of water and land management are only moderate (Haddeland et al., 2014).

4.6. Africa

Of the entire continent of Africa, only Southern Africa (SAF) has more than 50 stations fulfilling the data availability requirements. In the 1951–1990 period, low-flow indices (MIN and P10) and the annual median (P50) show significant decreasing regional trends. In 1961–2000 this weak drying pattern is reinforced, and all indices except P90 show significant negative trends, pointing at an overall decrease throughout the runoff distribution. In the 1971–2010 period, however, this pattern weakens and only MIN shows a significant decline.

A comprehensive assessment of trends in a region similar to SAF found more decreasing than increasing trends (Fanta et al., 2001), which is consistent with the findings of the present study. Overall, global climate models suggest that increasing global mean temperatures are associated with drying conditions in SAF (Greve et al., 2018), which might even be intensified through reduced flow rates triggered by human water and land management (Haddeland et al., 2014).

4.7. Oceania

In Oceania, only South Australia/New Zealand (SAU) has sufficient data coverage to warrant analysis. Data availability is not sufficient in the period 1951–1990. In 1961–2000, negligible change is observed, except weak inclination in MIN. However, the last period (1971–2010) shows strong and significant negative regional trends of all indices considered, that is, a strong and significant reduction of the entire flow distribution.

Previous assessments of changes in streamflow (Petrone et al., 2010; Zhang et al., 2016) and annual maximum floods (Ishak et al., 2013) have reported declining trends in southern Australia. In southeastern Australia, the first decade of the 21st century was particularly dry, sometimes referred to as the millennium drought (Kiem et al., 2016; Low et al., 2015; Zhang et al., 2016).

On average, the climate models of the CMIP5 ensemble indicate only a weak change in water availability in SAU with increasing global mean temperatures (Greve et al., 2018). However, other simulations suggest that both anthropogenic climate change and human water use may trigger a significant reduction of runoff in south eastern Australia (Haddeland et al., 2014), which is consistent with the observed changes.

5. Summary and Conclusions

To date, there has been low confidence and a lack of consistent evidence regarding sign and magnitude of trends in global river discharge during the twentieth century (Hartmann et al., 2013). Therefore, this study presents a comprehensive update of global-scale changes in indicators of mean and extreme streamflow taken from the GSIM archive (Do et al., 2018b; Gudmundsson et al., 2018b). To enable this global overview across all indicators, the focus of the analyses has been the significance and sign of change at the subcontinental scale, contrasting the common approach to solely report trend magnitudes for individual stations. In

contrast to regional studies tailored to specific indices and with varying methods, a key benefit of this study was the opportunity to consider multiple regions and multiple indices with a consistent method. The subsequent analysis highlights that streamflow trends have complex spatial patterns, preventing simple generalizations of regional changes to the global scale.

A striking result is that in most cases the sign of regional trends is consistent across all indices. This implies that the entire flow distribution is changing upward or downward in the respective regions, indicating generally wetter or drier conditions. In other words, increasing low flows are in most cases associated with increasing high flows (and vice versa), contradicting the common notion that flood and drought risk may increase simultaneously. Another feature of the results is that for some regions (West North America, East North America, and North-East Brazil) the sign of the trends has varied with respect to the considered period, suggesting low-frequency variability in the baseline climate signal and that care is needed in the interpretation of the associated change patterns.

Among all considered regions, South Europe/Mediterranean had the strongest signal with consistent negative trends in all indices throughout all considered time periods. Other regions with predominantly negative trends include Southern Africa, South Australia/New Zealand, and potentially South Asia. In addition, Northeastern Brazil experienced drying conditions for the last two time periods but had a consistent wetting trend for the first period. Consistent wetting trends were observed in Central North America, Southeastern South America, North Europe, and North Asia, although the trend weakens for the last period in Central North America and Southeast South America. Overall, these wetting trends are not equally visible in all regions and throughout all indices.

While the number of gauges used in this study is unprecedented, the conclusions in regions with less data are constrained (e.g., Asia) or muted (e.g., Africa), and further gains in data gathering would substantially improve confidence (Do et al., 2018b). Consequently, spatiotemporal coverage of the observations remains a limiting factor. Likewise, both the potentially uneven temporal distribution of available data in individual time series and regional differences in spatial coverage are impacting the results. Finally, the focus on regional trends can mask subregional features. Nevertheless, the presented results provide for an unprecedented view on streamflow trends around the world.

While this study has sought to interpret observed changes in the context of future climate projections (Greve et al., 2018) and model estimates of the impacts of human water and land management (Haddeland et al., 2014), it does not allow for a conclusive attribution of the observed changes to either of these factors. To this end, formal detection and attribution methods (Bindoff et al., 2013; Gudmundsson et al., 2017) are needed, which would allow for systematic testing of the hypothesis that both anthropogenic climate change and human water and land management are impacting renewable freshwater resources and hydrological extremes at the global scale. As end of the century projections of global water resources and hydrological extremes increase in number and sophistication, an appreciation for trends in observed indicators of mean and extreme streamflow provides a stronger basis for understanding future changes.

Acknowledgments
All streamflow indices considered are taken from the GSIM archive and are freely available from https://doi.org/10.1594/PANGAEA.887470 (Gudmundsson et al., 2018a). Mr. Hong Xuan Do receives financial support from the Australia Award Scholarship (AAS) and D. R. Stranks Traveling Fellowship. This work is partially funded by the ERC DROUGHT-HEAT project (contract 617518).

References

Adam, J. C., & Lettenmaier, D. P. (2008). Application of new precipitation and reconstructed streamflow products to streamflow trend attribution in northern Eurasia. *Journal of Climate*, *21*(8), 1807–1828.

Alkama, R., Decharme, B., Douville, H., & Ribes, A. (2011). Trends in global and basin-scale runoff over the late twentieth century: Methodological issues and sources of uncertainty. *Journal of Climate*, *24*(12), 3000–3014.

Berghuijs, W. R., Aalbers, E. E., Larsen, J. R., Trancoso, R., & Woods, R. A. (2017). Recent changes in extreme floods across multiple continents. *Environmental Research Letters*, *12*(11), 114,035–114,035.

Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., et al. (2013). Detection and attribution of climate change: From global to regional. In T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 867–952). Cambridge, UK: Cambridge University Press.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., et al. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Bouwer, L. M., Vermaat, J. E., & Aerts, J. C. J. H. (2006). Winter atmospheric circulation and river discharge in northwest Europe. *Geophysical Research Letters*, *33*, L06403. https://doi.org/10.1029/2005GL025548

Bouwer, L. M., Vermaat, J. E., & Aerts, J. C. J. H. (2008). Regional sensitivities of mean and peak river discharge to climate variability in Europe. *Journal of Geophysical Research*, *113*, D19103. https://doi.org/10.1029/2008JD010301

Burn, D. H. (2008). Climatic influences on streamflow timing in the headwaters of the Mackenzie River Basin. *Journal of Hydrology*, *352*(1–2), 225–238.

Burn, D. H., & Elnur, M. A. H. (2002). Detection of hydrologic trends and variability. *Journal of Hydrology*, *255*(1–4), 107–122.
Dai, A., Qian, T., Trenberth, K. E., & Milliman, J. D. (2009). Changes in continental freshwater discharge from 1948 to 2004. *Journal of Climate*, *22*(10), 2773–2792.
Dai, A., & Trenberth, K. E. (2002). Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *Journal of Hydrometeorology*, *3*(6), 660–687.
Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018a). The global streamflow indices and metadata archive—Part 1: Station catalog and catchment boundary, edited, PANGAEA.
Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018b). The Global Streamflow Indices and Metadata archive (GSIM)—Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data*, *10*(2), 765–785.
Do, H. X., Westra, S., & Leonard, M. (2017). A global-scale investigation of trends in annual maximum streamflow. *Journal of Hydrology*, *552*, 28–43.
Douglas, E. M., Vogel, R. M., & Kroll, C. N. (2000). Trends in floods and low flows in the United States: Impact of spatial correlation. *Journal of Hydrology*, *240*(1–2), 90–105.
ECA & D Project Team, and K. Royal Netherlands Meteorological Institute (2013). Algorithm Theoretical Basis Document (ATBD). Rep. Retrieved from https://www.ecad.eu/documents/atbd.pdf
Fanta, B., Zaake, B. T., & Kachroo, R. K. (2001). A study of variability of annual river flow of the southern African region. *Hydrological Sciences Journal*, *46*(4), 513–524.
Fekete, B. M., Looser, U., Pietroniro, A., & Robarts, R. D. (2012). Rationale for monitoring discharge on the ground. *Journal of Hydrometeorology*, *13*(6), 1977–1986.
Fekete, B. M., Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E., & Zhulidov, A. V. (2015). Time for in situ renaissance. *Science*, *349*(6249), 685–686.
Genta, J. L., Perez-Iribarren, G., & Mechoso, C. R. (1998). A recent increasing trend in the streamflow of rivers in southeastern South America. *Journal of Climate*, *11*(11), 2858–2862.
Greve, P., Gudmundsson, L., & Seneviratne, S. I. (2018). Regional scaling of annual mean precipitation and water availability with global temperature change. *Earth System Dynamics*, *9*(1), 227–240.
Gudmundsson, L., Do, H. X., Leonard, M., & Westra, S. (2018a). The Global Streamflow Indices and Metadata archive (GSIM)—Part 2: Time series indices and homogeneity assessment, edited, PANGAEA.
Gudmundsson, L., Do, H. X., Leonard, M., & Westra, S. (2018b). The Global Streamflow Indices and Metadata archive (GSIM)—Part 2: Quality control, time-series indices and homogeneity assessment. *Earth System Science Data*, *10*(2), 787–804.
Gudmundsson, L., Seneviratne, S. I., & Zhang, X. (2017). Anthropogenic climate change detected in European renewable freshwater resources. *Nature Climate Change*, *7*, 813–816.
Haddeland, I., Heinke, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., et al. (2014). Global water resources affected by human interventions and climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(9), 3251–3256. https://doi.org/10.1073/pnas.1222475110
Hannaford, J., Buys, G., Stahl, K., & Tallaksen, L. M. (2013). The influence of decadal-scale variability on trends in long European streamflow records. *Hydrology and Earth System Sciences*, *17*(7), 2717–2733.
Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., et al. (2011). Large-scale river flow archives: Importance, current status and future needs. *Hydrological Processes*, *25*(7), 1191–1200.
Hartmann, D. L., Klein Tank, A. M. G., Rusicucci, M., Alexander, L. V., Broenniman, B., Charabi, Y., et al. (2013). Observations: Atmosphere and surface. In T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 159–254). Cambridge, UK: Cambridge University Press.
Helsel, D. R., & Frans, L. M. (2006). Regional Kendall test for trend. *Environmental Science & Technology*, *40*(13), 4066–4073.
Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., et al. (2013). Global flood risk under climate change. *Nature Climate Change*. https://doi.org/10.1038/nclimate1911
Hisdal, H., Stahl, K., Tallaksen, L. M., & Demuth, S. (2001). Have streamflow droughts in Europe become more severe or frequent? *International Journal of Climatology*, *21*, 317–333.
Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., et al. (2017). Climate-driven variability in the occurrence of major floods across North America and Europe. *Journal of Hydrology*, *552*, 704–717. https://doi.org/10.1016/j.jhydrol.2017.07.027
Ishak, E. H., Rahman, A., Westra, S., Sharma, A., & Kuczera, G. (2013). Evaluating the non-stationarity of Australian annual maximum flood. *Journal of Hydrology*, *494*, 134–145.
Jaramillo, F., & Destouni, G. (2015). Local flow regulation and irrigation raise global human water consumption and footprint. *Science*, *350*(6265), 1248–1251.
Kiem, A. S., Johnson, F., Westra, S., van Dijk, A., Evans, J. P., O'Donnell, A., et al. (2016). Natural hazards in Australia: Droughts. *Climatic Change*, *139*(1), 37–54. https://doi.org/10.1007/s10584-016-1798-7
Kingston, D. G., Fleig, A. K., Tallaksen, L. M., & Hannah, D. M. (2012). Ocean-atmosphere forcing of summer streamflow drought in Great Britain. *Journal of Hydrometeorology*, *14*(1), 331–344.
Kuhnel, I., McMahon, T. A., Finlayson, B. L., Haines, A., Whetton, P. H., & Gibson, T. T. (1990). Climatic influences on streamflow variability: A comparison between southeastern Australia and southeastern United States of America. *Water Resources Research*, *26*(10), 2483–2496.
Kundzewicz, Z. W., Graczyk, D., Maurer, T., Pĭnskwar, I., Radziejewski, M., Svensson, C., & Szwed, M. g. (2005). Trend detection in river flow series: 1. Annual maximum flow. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, *50*(5), 797–810.
Labat, D., Goddéris, Y., Probst, J. L., & Guyot, J. L. (2004). Evidence for global runoff increase related to climate warming. *Advances in Water Resources*, *27*(6), 631–642.
Lettenmaier, D. P., Wood, E. F., & Wallis, J. R. (1994). Hydro-climatological trends in the continental United States, 1948-88. *Journal of Climate*, *7*(4), 586–607.
Lins, H. F., & Slack, J. R. (1999). Streamflow trends in the United States. *Geophysical Research Letters*, *26*(2), 227–230. https://doi.org/10.1029/1998GL900291
Low, K. G., Grant, S. B., Hamilton, A. J., Gan, K., Saphores, J.-D., Arora, M., & Feldman, D. L. (2015). Fighting drought with innovation: Melbourne's response to the Millennium Drought in Southeast Australia. *Wiley Interdisciplinary Reviews Water*, *2*(4), 315–328.
MacDonald, G. M., Kremenetski, K. V., Smith, L. C., & Hidalgo, H. G. (2007). Recent Eurasian river discharge to the Arctic Ocean in the context of longer-term dendrohydrological records. *Journal of Geophysical Research*, *112*, G04S50. https://doi.org/10.1029/2006JG000333

Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, *5*(3), 250–254.

Marengo, J. A., Tomasella, J., & Uvo, C. R. (1998). Trends in streamflow and rainfall in tropical South America: Amazonia, eastern Brazil, and northwestern Peru. *Journal of Geophysical Research*, *103*(D2), 1775–1783. https://doi.org/10.1029/97JD02551

McCabe, G. J., & Wolock, D. M. (2002). A step increase in streamflow in the conterminous United States. *Geophysical Research Letters*, *29*(24), 2185. https://doi.org/10.1029/2002GL015999

McClelland, J. W., Déry, S. J., Peterson, B. J., Holmes, R. M., & Wood, E. F. (2006). A pan-Arctic evaluation of changes in river discharge during the latter half of the 20th century. *Geophysical Research Letters*, *33*, L06715. https://doi.org/10.1029/2006GL025753

Milliman, J. D., Farnsworth, K. L., Jones, P. D., Xu, K. H., & Smith, L. C. (2008). Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951--2000. *Global and Planetary Change*, *62*(3–4), 187–194.

Milly, P. C. D., Dunne, K. A., & Vecchia, A. V. (2005). Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, *438*(7066), 347–350.

Mondal, A., & Mujumdar, P. P. (2012). On the basin-scale detection and attribution of human-induced climate change in monsoon precipitation and streamflow. *Water Resources Research*, *48*, W10520. https://doi.org/10.1029/2011WR011468

Pasquini, A. I., & Depetris, P. J. (2007). Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview. *Journal of Hydrology*, *333*(2), 385–399.

Petrone, K. C., Hughes, J. D., van Niel, T. G., & Silberstein, R. P. (2010). Streamflow decline in southwestern Australia, 1950-2008. *Geophysical Research Letters*, *37*, L11401. https://doi.org/10.1029/2010GL043102

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., et al. (2014). Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(9), 3262–3267. https://doi.org/10.1073/pnas.1222473110

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., et al. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(9), 3245–3250. https://doi.org/10.1073/pnas.1222460110

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's Tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi et al. (Eds.), *Managing the risks of extreme events and disasters to advance climate change adaptation, A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC)* (pp. 109–230). Cambridge, UK and New York: Cambridge University Press. Retrieved from https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3_FINAL-1.pdf

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., et al. (2010). Streamflow trends in Europe: Evidence from a dataset of near-natural catchments. *Hydrology and Earth System Sciences*, *14*(12), 2367–2382.

Stahl, K., Tallaksen, L. M., Hannaford, J., & van Lanen, H. A. J. (2012). Filling the white space on maps of European runoff trends: Estimates from a multi-model ensemble. *Hydrology and Earth System Sciences*, *16*(7), 2035–2047.

Svensson, C., Kundzewicz, W. Z., & Maurer, T. (2005). Trend detection in river flow series: 2. Flood and low-flow index series. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, *50*(5), 811–824.

Tananaev, N. I., Makarieva, O. M., & Lebedeva, L. S. (2016). Trends in annual and extreme flows in the Lena River basin, northern Eurasia. *Geophysical Research Letters*, *43*, 10,764–10,772. https://doi.org/10.1002/2016GL070796

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2011). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Tootle, G. A., & Piechota, T. C. (2006). Relationships between Pacific and Atlantic Ocean sea surface temperatures and U.S. streamflow variability. *Water Resources Research*, *42*, W07411. https://doi.org/10.1029/2005WR004184

Tootle, G. A., Piechota, T. C., & Singh, A. (2005). Coupled oceanic-atmospheric variability and U.S. streamflow. *Water Resources Research*, *41*, W12408. https://doi.org/10.1029/2005WR004381

Verdon, D. C., Wyatt, A. M., Kiem, A. S., & Franks, S. W. (2004). Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resources Research*, *40*, W10201. https://doi.org/10.1029/2004WR003234

Wanders, N., & Wada, Y. (2015). Decadal predictability of river discharge with climate oscillations over the 20th and early 21st century. *Geophysical Research Letters*, *42*, 10,689–10,695. https://doi.org/10.1002/2015GL066929

Ward, P. J., Beets, W., Bouwer, L. M., Aerts, J. C. J. H., & Renssen, H. (2010). Sensitivity of river discharge to ENSO. *Geophysical Research Letters*, *37*, L12402. https://doi.org/10.1029/2010GL043215

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., & Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(9), 3228–3232.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (p. 676). Oxford, Amsterdam, and Waltham, MA: Academic Press.

Zhang, X. S., Amirthanathan, G. E., Bari, M. A., Laugesen, R. M., Shin, D., Kent, D. M., et al. (2016). How streamflow has changed across Australia since the 1950s: Evidence from the network of hydrologic reference stations. *Hydrology and Earth System Sciences*, *20*(9), 3947–3965.

Supplementary Article 3

Addor, N., Alvarez-Garreto, C., Coxon, G., Do, H.X., Gudmundsson, L., Fowler, K., and Mendoza, P. (2019). Large-sample hydrology: recent progress, guidelines for new datasets and outstanding challenges (for submission 03/2019 to Hydrological Sciences Journal special issue on "Hydrological data: opportunities and barriers").

Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges

Journal:	*Hydrological Sciences Journal*
Manuscript ID	HSJ-2019-0176
Manuscript Type:	Special Issue Paper
Date Submitted by the Author:	31-Mar-2019
Complete List of Authors:	Addor, Nans; University of East Anglia, Do, Hong; University of Adelaide School of Civil Environmental and Mining Engineering; Nong Lam University, Faculty of Environment and Natural Resources Alvarez-Garreton, Camila; Universidad Austral de Chile, Instituto de Conservación, Biodiversidad y Territorio; Center for Climate and Resilience Research Coxon, Gemma ; University of Bristol Fowler, Keirnan; University of Melbourne, Department of Infrastructure Engineering Mendoza, Pablo; Universidad de Chile, Department of Civil Engineering; Universidad de Chile, Advanced Mining Technology Center
Keywords:	streamflow records, data standardisation, reproducibility of hydrological experiments, data uncertainties, human interventions, cloud computing

Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges

Manuscript submitted on 31/03/2019 to the Hydrological Sciences Journal, special issue on "Hydrological data: opportunities and barriers"

Nans Addor[1,*], Hong Xuan Do[2,3], Camila Alvarez-Garreton[4,5], Gemma Coxon[6], Keirnan Fowler[7], Pablo A. Mendoza[8,9]

[1] Climatic Research Unit, School of Environmental Sciences, University of East Anglia, UK
[2] School of Civil, Environmental and Mining Engineering, University of Adelaide, Australia
[3] Faculty of Environment and Natural Resources, Nong Lam University, Vietnam
[4] Instituto de Conservación, Biodiversidad y Territorio, Universidad Austral de Chile, Valdivia, Chile
[5] Center for Climate and Resilience Research, Chile
[6] School of Geographical Sciences, University of Bristol, United Kingdom
[7] Department of Infrastructure Engineering, University of Melbourne, Australia
[8] Department of Civil Engineering, Universidad de Chile, Chile
[9] Advanced Mining Technology Center, Universidad de Chile, Chile
*Correspondence: N.Addor@uea.ac.uk

Keywords: streamflow records, data standardisation, reproducibility of hydrological experiments, data uncertainties, human interventions, cloud computing.

Highlight fields

- We review currently available data sets for large-sample hydrology (LSH) and highlight the breadth of hydrological studies they underpin
- We stress that LSH datasets are so far limited by their lack of comparability, uncertainty estimates and characterisation of human impacts
- We propose guidelines for new LSH datasets and coordinated actions to overcome these barriers

Abstract

Large-sample hydrology (LSH) relies on data from large samples (tens to thousands) of catchments to go beyond individual case studies and derive robust conclusions on hydrological processes and models. Several LSH datasets have recently been released, covering a wide range of regions and relying on increasingly diverse data sources to characterise catchment behaviour. While these datasets offer novel opportunities, they are also limited by their lack of comparability, uncertainty

estimates and characterisation of human impacts. This article i) underscores the key role of LSH datasets in hydrological studies, ii) provides a review of currently available LSH datasets, iii) highlights current limitations of LSH datasets and iv) proposes guidelines and coordinated actions to overcome these limitations. These guidelines and actions aim to standardise and automatise the maintenance of LSH datasets worldwide, and to enhance the reproducibility and comparability of hydrological studies.

1. Introduction: from comparative hydrology to large-sample hydrology

Large-sample hydrology (LSH) makes use of datasets involving large samples of catchments to derive robust conclusions on hydrological processes and models. LSH finds its roots in the field of comparative hydrology (Kovács, 1984; Falkenmark and Chapman, 1989), whose foundations were set in the framework of International Hydrological Programme, launched by UNESCO in 1975. At that time, a practical objective was to determine to what extent available hydrological techniques and models, which were derived mostly for temperate regions of Europe and North America, could be transferred to and applied in other regions. The more general motivation of comparative hydrology is to learn from hydrological similarities and differences between places around the world, and interpret these in terms of underlying climate-landscape-human controls (e.g., Finlayson et al., 1986; McMahon, 1982; Peel et al., 2001, 2004; Sivapalan, 2009; Troch et al. 2009; Thomson et al., 2011).

LSH follows similar objectives but puts a stronger emphasis on the need to establish robust principles by leveraging large samples of observations, which led to the name of “large-sample hydrology”. Andréassian et al. (2006a) underscore that model intercomparisons should be based on a significant number of catchments to deliver robust conclusions that are not the result of chance. Similarly, Gupta et al. (2014) insist that general hydrological principles should be derived from statistically significant relationships, which are unobtainable with data from only a few catchments. This makes LSH a branch, rather than a replacement of comparative hydrology, and thus several comparative hydrology investigations can also be classified as LSH research (e.g., Singh et al., 2014).

Alongside large-sample hydrology, large-scale hydrology has become established (Cloke and Hannah, 2011; Wood et al., 2011; Bierkens, 2015). These two fields are complementary, as they both provide generalisable knowledge on the terrestrial water cycle across a range of hydroclimatic conditions. A notable difference between them lies in the scale and spatial continuity of the area covered. A large sample of catchments can cover a vast area, but this area is made of separate basins. In contrast, large-scale hydrology explores “spatial scales greater than a single river basin all the way up to the entire planet” to use the definition of Cloke and Hannah

(2011). Further, while streamflow measurements are a cornerstone of catchment hydrology and LSH, at larger spatial scales, the focus is traditionally on other fluxes (e.g., evapotranspiration) and state variables (e.g, soil moisture). The gap between these two fields is however quickly reducing, with the development of gridded streamflow observations (Fekete et al., 2002; Gudmundsson and Seneviratne, 2016; Ghiggi et al., 2019), ever larger domains covered by rainfall-runoff models (Beck et al., 2016), the ever finer resolution of large-scale models (Wood et al., 2011; Bierkens et al., 2015) and following efforts to evaluate the influence of catchment-scale processes on the performances of large-scale models (Fang et al., 2017; Kauffeldt et al., 2016; Veldkamp et al., 2018; Zaherpour et al., 2018) and to include streamflow simulations from macroscale models in LSH investigations (e.g., Rakovec et al., 2016; Zink et al., 2017).

In this paper, we focus on LSH and, more specifically, on datasets providing streamflow data for a large number of catchments. Such datasets form the foundation of a wide range of hydrological studies dedicated to catchment classification (e.g., Sawicz et al., 2011; Kuentz et al. 2017; Knoben et al., 2018), extreme events (e.g., Tijdeman et al., 2016; Berghuijs et al., 2017; Blöschl et al., 2017; Do et al., 2017; Mallakpour and Villarini, 2015; Gudmundsson et al., 2019), terrestrial water storage (e.g., Zhang et al., 2017), data and model uncertainties (e.g., McMillan et al, 2012; Coxon et al, 2015; Beck et al. 2017), hydrological model evaluation and benchmarking (e.g., Gudmundsson et al., 2012; Coron et al., 2012; Coxon et al., 2013; Fowler et al., 2016; McMillan et al., 2016a; Newman et al., 2017; Seibert et al., 2018, Kratzert et al., 2018), parameter estimation of hydrological models (e.g., Oudin et al., 2010; Beck et al., 2016; Rakovec et al., 2016; Hirpa et al, 2018), regionalisation using machine learning algorithms (Beck et al., 2015; Addor et al., 2018; Barbarossa et al., 2018), human impacts on hydrology (e.g., Alvarez-Garreton et al., 2018; Tijdeman et al., 2018a, 2018b), streamflow forecasting (e.g., Harrigan et al., 2018; Slater and Villarini, 2018) and climate change impacts assessments (e.g., Melsen et al., 2018). LSH datasets underpinn key advances in hydrological sciences and are fundamental to major community-wide efforts, in particular to the Prediction in Ungauged Basins (PUB, Hrachowitz et al., 2013) and Panta Rhei (Montanari et al, 2013; McMillan et al., 2016b) initiatives of the International Association of Hydrological Sciences (IAHS).

The diversity and content of LSH datasets is expanding rapidly. Gupta et al. (2014) highlighted several datasets potentially useful for LSH applications and since then, several datasets dedicated to LSH have been published. They cover a far greater number of catchments, hydroclimatic regions and catchment attributes than what was available just a few years ago. In Section 2, we provide a snapshot of this development and give an overview of LSH datasets currently available. These recent advances and the opportunities they offer are remarkable, yet, as creators and users of LSH datasets, we argue that it is now crucial to better coordinate the production and exchange of LSH datasets worldwide. For this Hydrological Sciences Journal

special issue on "Hydrological data: opportunities and barriers", we identified four LSH challenges that require immediate attention: (i) the difficulties of inter-database comparison, (ii) the lack of uncertainty estimates, (iii) the insufficient representation of human interventions in LSH datasets, and (iv) the still limited accessibility of streamflow observations. These challenges are discussed in Section 3. We then list simple, concrete actions (Section 4) and outline coordinated efforts (Section 5) to overcome these barriers. Conclusions are presented in Section 6.

2. Recent progress in the development of LSH datasets

In this section, we review LSH datasets currently available, and focus on LSH datasets fulfilling two criteria, referred to below as "minimum requirements": (i) the dataset must contain streamflow observations and (ii) basic identifiers for each stream-gauge (i.e., name, catchment area, gauge coordinates) must be included. We did not set a specific number of catchments to define a sample as "large", as the needs of each study are unique. For instance, tens of carefully selected catchments can enable insightful regional comparisons (e.g., Bennett et al., 2018; Burn and Whitfield, 2018; Fowler et al., 2018), while one may argue that thousands of catchments are needed for global scale investigations (e.g., Beck et al., 2015; Do et al., 2017; Gudmundsson et al., 2019; van Dijk et al., 2013). In addition, this paper focuses on datasets available in digital form with relative ease of access. It does not cover individual national water archives, the classical data source resulting from national-scale streamflow monitoring, as some of them are only maintained in paper form or subject to strict data-distribution policies. However, these national archives form the basis of the LSH datasets described below.

2.1. Data available through LSH datasets

The nature of the data covered by LSH datasets varies significantly from one dataset to the next. To facilitate the navigation and selection of LSH datasets by potential users, here we classify these data into three categories: (i) streamflow observations, (ii) hydrometeorological time series and (iii) landscape and hydroclimatic attributes.

Streamflow observations is a category on its own, since we make their availability a minimum requirement for the dataset to be consider here (Table 1). Some LSH datasets complement streamflow observations with other hydrometeorological time series, such as precipitation and temperature. Further, indices characterising the landscape of the catchments, for instance their land cover or soil, are included in some datasets. We note that the availability of hydrometeorological time series and catchment landscape attributes varies strongly among LSH data sets. The wealth of available spatial data products (e.g., different remotely sensed vegetation products for land cover or sub-daily meteorological variables) means that LSH datasets creators only select a subset of these available datasets. As a result, different LSH datasets are best adapted to different research pursuits. For example, datasets

including atmospheric forcing time series for each catchment (Addor et al., 2017; Arsenault et al., 2016; Schaake et al., 2006; Alvarez-Garreton et al., 2018) are well suited for hydrological modelling, whereas data products including catchment attributes representing human presence or irrigation (Do et al., 2018a; Gudmundsson et al., 2018a) are more adapted for detection and attribution studies.

In addition, some datasets provide metadata and uncertainty estimates. For example, catchment boundaries may be provided with quality flags, and time series may be subject to a homogeneity assessment to produce uncertainty estimates (Do et al., 2018a, Gudmundsson et al., 2018a). Other data sets derived meteorological time series using several data-products to reflect forcing uncertainty (Alvarez-Garreton et al., 2018; Newman et al. (2015a).

2.2. LSH datasets currently available

Table 2 provides an overview of eleven key LSH datasets. These datasets cover different parts of the world and include basins from a single country to the entire globe. The access to these datasets is unrestricted for scientific purposes. However, the licensing policies vary with some datasets being fully available in the public domain, while others requiring data requests in written form.

At the global scale, the Global Runoff Data Base (GRDB) is arguably the main dataset used for streamflow investigations, including LSH studies. This database is maintained by the Global Runoff Data Centre (GRDC), which operates under the auspices of the World Meteorological Organization (WMO) since 1988, and holds records of daily and monthly streamflow across more than 9,000 stations globally (GRDC, 2015). This global initiative is supported voluntarily by national authorities and thus, data contributions depend on the capacity of corresponding agencies. As a result, some countries are sparsely represented in GRDB, even though data of reasonable quality are available (e.g., most stations in Asia have not been updated since the 1990s; GRDC 2015). To facilitate access to streamflow data from stations across the world, the Global Streamflow Indices and Metadata (GSIM) archive was recently produced (Do et al., 2018a; Gudmundsson et al., 2018a). GSIM is an expansion of GRDB, which was produced by collating streamflow observations from 11 other publicly available databases (including three LSH datasets also described in Table 2) and publishing standardized metadata relevant to LSH research (Do et al., 2018b). To make hydrological information publicly available, even when raw data cannot be redistributed, GSIM contains time series of streamflow indices at different temporal resolutions (i.e., monthly/seasonal/yearly time-step) derived from raw daily records (Gudmundsson et al., 2018b).

At the continental scale, the European Water Archive (EWA) is one of the most comprehensive streamflow time-series archives with records of more than 3000 river gauging stations contributed by 29 European national water agencies. EWA is now

hosted by GRDC and thus data is available under the GRDC data policy. This data archive, however, has been frozen since 2014 until GRDC is permitted to integrate EWA stations into GRDB. EWA streamflow records were recently combined with GRDB stations and the European catchments from the Hydrological Predictions for the Environment model to improve the characterisation and understanding of hydrologic availability across Europe through catchment classifications (Kuentz et al., 2017). Another long-standing LSH dataset is the data archive of the Model Parameter Estimation Experiment project (MOPEX; Duan et al., 2006). In addition to hydrometeorological observations, MOPEX provides attributes for catchments representing different hydroclimatic conditions. MOPEX includes data for 438 catchments in the United States and was one of the main data sources underpinning the PUB decade (Andreassian et al., 2006b). However, MOPEX hydrometeorological time series stop in 2003 and MOPEX is not anymore updated.

At the national scale, several datasets have been developed with an approach similar to that of MOPEX. The Catchment Attributes and MEteorology for Large-sample Studies dataset (CAMELS; Newman et al., 2015a; Addor et al., 2017) uses recent datasets to provide up-to-date hydrometeorological variables and a variety of landscape attributes for 671 catchments across the contiguous United States. CAMELS also includes detailed descriptions of the methods used to derive catchment attributes and a discussions of several of data-source caveats. A similar approach was used to produce the CAMELS-Chile dataset (CAMELS-CL; Alvarez-Garreton et al., 2018), which provides an overview of regional variations in hydroclimatic conditions over Chile and an assessment of human interventions to streamflow regime across 516 catchments. Meteorological and hydrologic data for 698 catchments in Canada are available through the CANOPEX database (Arsenault et al., 2016).

3. Limitations of current LSH datasets

In this section, we highlight four typical limitations of LSH datasets: i) the lack of common standards impedes the comparison of basins from different datasets, ii) the lack of uncertainty estimates prevents users from assessing data reliability, iii) the extent of human interventions is rarely characterised, and iv) data accessibility is still limited.

3.1 The lack of common standards impedes the comparison of basins from different datasets

Comparative hydrology is only possible if the data from different catchments are consistently processed, and thus can be compared. Although the comparison of catchments from the same LSH dataset is usually straightforward, comparisons across LSH datasets is challenging because different naming conventions, data sources and ways of calculating the same variable are used from one dataset to the

next. This issue is part of the wider challenge of using common standards and protocols when producing and processing environmental data (Horsburgh et al., 2008; Ceola et al., 2015), and it critically limits our ability to combine and learn from several LSH datasets.

3.2 The lack of metadata and uncertainty estimates prevents users from assessing data reliability

When using data from many catchments, assessing data errors is key, as they can bias comparisons between catchments. Yet, there is still a clear lack of uncertainty estimates accompanying LSH datasets. Uncertainties in the atmospheric forcing are receiving the most attention and are increasingly characterised by relying on several datasets (e.g., Newman et al., 2015a; Alvarez-Garreton et al., 2018). In contrast, uncertainties in the catchment attributes (e.g., land cover, soil characteristics) are rarely quantified, or even acknowledged, in LSH datasets.

Streamflow uncertainty estimates and metadata on gauge information are also rarely available in LSH datasets, although the limitations and uncertainties of streamflow time series are well known (e.g, McMillan et al, 2012). Streamflow metadata are often not available due to management practices of data providers (Hannah et al., 2011), the loss of metadata during data transfers from providers to international data archives or poor upkeep of this information (Gudmundsson et al., 2018a). Further, even when metadata are available, robustly assessing streamflow uncertainties across large samples of catchments remains a challenge, as different methods are recommended for different gauge types (Kiang et al., 2018).

3.3 The extent of human interventions is rarely characterised

LSH datasets have historically focused on physical attributes, making use of the wealth of data currently available to characterise hydrological behaviour (Tables 1 and 2). In comparison, human interventions are still poorly characterised in LSH datasets, although human alterations have large impacts on the natural water cycle (e.g., Vorosmarty et al., 2000, Hanasaki et al., 2006). These impacts may be comparable to climate change effects at the regional scale (Ferguson and Maxwell, 2012) and threaten sustainability at the global scale (Jaramillo and Destouni, 2015). For example, increased reservoir storage at the catchment scale not only affects runoff seasonality, but also the frequency of low/high flow events observed at the catchment outlet (e.g., Wehren et al., 2010), and changes in land cover influence the distribution of streamflow, specifically baseflow volumes and flashiness of runoff (e.g. Vertessy 2000; Brown et al., 2005). Consequently, providing information on such alterations is critical to assess the magnitude of human impacts on catchment-scale hydrological behavior (e.g., Alvarez-Garreton et al., 2018) and to incorporate human interventions in hydrological model (e.g., Liu et al. 2017, Veldkamp et al. 2018) and thus provide reliable hydrological simulations in an increasingly built-up environment.

3.4 LSH datasets are rarely FAIR - Findable, Accessible, Interoperable and Reusable

To advance LSH, progress is needed to make LSH datasets more FAIR (Findable, Accessible, Interoperable and Reusable, see Wilkinson et al., 2016 and the Open Data Charter, 2015). Currently, many digitised datasets are stored in local repositories or the data portals are unknown to data users (not 'Findable'). The accessibility of streamflow data is still limited for many regions of the world (not 'Accessible') biasing LSH studies towards countries with greater accessibility. LSH datasets are hosted in different locations with a range of different upkeeping practices (not 'Interoperable'). The license of many streamflow records does not allow users to share streamflow records in their possession (not 'Reusable').

Differences in the availability of streamflow records worldwide are highlighted in Figure 1. The map indicates the varying temporal coverage across the globe, with stations in North America and Europe generally having the longest records. Importantly, 'white space' still dominates in many regions of the world, as also shown in other studies (e.g., Barbarossa et al., 2018). In some cases, this can be attributed to the lack of stations, in particular in extreme environments. Although in several regions data streamflow records do exist, they are not accessible because (i) data are not available in digitised form, (ii) digitised data are hosted in a local repository and data authorities do not have available resource to process data requests, (iii) data are not made available or are subject to payable fees, and (iv) the one-station-at-a-time downloading process (mouse and keyboard interactions required) hampers data retrieval for LSH studies.

4. Guidelines for the production of LSH datasets

To overcome the limitations outlined in Section 3, we propose six simple guidelines to support the creation of future LSH datasets (presented in this Section) and coordinated actions (presented in Section 5). The limitations, guidelines and actions are summarised graphically in Figure 2.

The six guidelines outlined here are simple to follow and will significantly improve the value and usability of future datasets. We consider them as minimum requirements to be satisfied by new LSH datasets, and hence suggest that they are checked by both LSH dataset creators and by reviewers of papers introducing new LSH datasets.

1. Provide basic data for each basin

Streamflow observations remain the cornerstone of LSH, and thus all new LSH datasets should make these records available. For streamflow records subject to strict redistribution data policy, releasing streamflow indices at different temporal

resolutions is an alternative (e.g., Do et al., 2018a, Gudmundsson et al., 2018a). Station metadata should at least include the upstream drainage area, river name and geographical coordinates of each stream-gauge. Providing the catchment boundary associated with each stream-gauge should also be prioritised, so that users can derive additional attributes or time series from global or regional data products. Using the same digital elevation data source for all the basins is recommended, HydroSHEDS (http://www.hydrosheds.org) and Viewfinder (http://viewfinderpanoramas.org) being popular choices at the global scale.

2. Follow established standards when naming variables

The observance of common standards, including the use of a controlled vocabulary, is essential to ensure the consistency and comparability of environmental datasets (e.g., Horsburgh et al., 2009; Vitolo et al, 2015.; Moine et al., 2014). Consistent variable names across LSH datasets should be used to make new datasets easier to utilise by the community and to facilitate inter-dataset comparisons. This is in addition to metadata, which describe the methods and data sources used to compute each variable.

Standards already exist for climate variables (Climate and Forecast Community Metadata Standard, http://cfconventions.org) and hydrological variables (http://www.waterml2.org). Although these standards form the bases of variable naming, they only cover a fraction of all the variables relevant to LSH. Hence, our recommendation is build on these standards, consider naming decisions made in other LSH datasets and improve them, with the goal to create a set of variable names that can be used across LSH datasets.

3. Use publicly available code for data processing

To improve transparency and reproducibility, the code used for the creation of LSH datasets should be publicly available, either by using already-published code (e.g., packages) or making the code used available (e.g., on GitHub). Several packages and libraries already exist to compute key attributes (especially climate indices and hydrological signatures) in different languages (see https://github.com/ropensci/hydrology and Slater and al., 2019 for R). Given that hydrological signatures can be particularly sensitive to their formulation, as shown for instance by Stoelzle et al. (2013) for recession coefficients, using publicly available code is essential.

Ideally, the shared code should cover more than the computation of climatic indices and hydrological signatures. It should, for instance, also include scripts to create catchment-averages from gridded products, and algorithms performing quality assurance tests of streamflow data. The goal is to create a library of scripts to perform standards LSH dataset creation tasks, thereby increasing the transparency

and comparability. For instance, the scripts used to produce the CAMELS attribute dataset (Addor et al., 2017) are publicly available (https://github.com/naddor/camels) and have been used to produce the CAMELS-CL dataset (Alvarez-Garreton et al, 2018).

4. Provide uncertainty estimates for time series and catchment attributes

To allow users to assess the reliability of a LSH dataset, their quality should be evaluated and provided as metadata alongside the dataset. Quality flags from data providers and simple numerical screening techniques can be used to develop quality assurance (QA) methods (see for example Gudmundsson et al, 2018, for flow QA procedures and Blenkinsop et al, 2016 and Lewis et al, 2018 for rainfall QA procedures). This should be developed in cooperation with hydrometric agencies who often employ QA procedures before the data is released.

The uncertainty of data products used when producing a LSH dataset should be assessed. One opportunity for large sample hydrology is to construct multiple estimates of a given variable using different products or formulations. This is already evident in many of the LSH datasets highlighted in Section 2, for example generating daily estimates of precipitation and potential evapotranspiration from multiple products (e.g., Alvarez-Garreton et al, 2018), and is becoming more viable with the increasing availability of continental/global products (e.g., see for example Beck et al, 2017).

5. Include descriptors of water administration systems

Water administration descriptors should be included in LSH datasets. Ideally, the following attributes should be provided at the most detailed spatial and temporal scales possible: (i) usage type (e.g., consumption, irrigation, hydropower, groundwater recharge), (ii) location, (iii) allocated volume, and (iv) timing of extraction. The first attribute indicates whether water returns to the rivers are expected, and hence should be completed by additional information for attributes (ii)-(iv). Other information used as a proxy for human water use such as catchment population, percentage of urban and agricultural land use, and the presence of dams (http://globaldamwatch.org) are also valuable, particularly in regions where water usage data are not available.

6. Assess and increase the dataset FAIRness

For new LSH datasets to be findable, they should be documented in open-access, peer-reviewed journals and indexed via a DOI within publication databases. Within-agency technical reports are not sufficient to ensure findability. To be accessible, datasets should be downloadable from the internet at no cost and provide the option to download the entire dataset at once (in addition to site-by-site download). To be

interoperable, the meaning of the data should be unambiguous regardless of the context. To be reusable, datasets need a licence allowing users to use, share, and build upon the existing dataset, encouraging collaboration and extension of existing datasets. We recommend to data owners to assess and increase the FAIRness of their dataset by using this online tool by the Australian National Data Service and partners (https://www.ands-nectar-rds.org.au/fair-tool, ANDS et al., 2017).

5. Outlook: grand challenges and priorities for LSH

In this section, we discuss tasks that go beyond what can be expected from an individual LSH study, and require coordinated efforts from the LSH community, and in some cases, the wider community. We deliberately kept their formulation general and not prescriptive to stimulate discussion in the community. These four challenges are ranked based on their spatial scale: from the challenges requiring a global strategy to those relying on efforts at the national and regional scale.

5.1 Facilitate the creation and increase the comparability of LSH datasets by moving their production to the cloud

We propose that the production of LSH data should be progressively moved to the cloud. Currently, LSH creators download different versions of various data products and process them using different scripts. As an alternative, the relevant global datasets should be available in a single place in the cloud, together with scripts necessary to process them. Users would upload shapefiles of their catchments and the extraction of hydrometeorological time series and catchment attributes would happen online. This would i) improve inter-dataset comparability as both data products and scripts would be the same, ii) facilitate the production of time series and attributes for new catchments, iii) enable the simultaneous update of LSH datasets, for instance when a new data product becomes available or covers a longer period. Such a system, accessible and maintained by the community instead of a few individuals, would increase the perennity of LSH datasets, i.e., make them easier to produce and maintain in the mid- to long-term.

We acknowledge that, because of licensing restrictions, some data, such as streamflow data, cannot be uploaded to the cloud. There are, however, a growing number of open, global datasets covering a variety of variables relevant for LSH, which can be processed online (see for instance data products involved in Addor et al., 2017; Beck et al., 2017; Nijzink et al., 2018; Rakovec et al., 2016). Furthermore, diverse data products, in particular remotely sensed data, are already available on cloud computing portals such as Google Earth Engine (https://developers.google.com/earth-engine/datasets/catalog/) and Amazon Web Services (https://registry.opendata.aws). Alternatively, the data processing may be arranged on a non-commercial data-sharing platform, such as Hydroshare (https://www.hydroshare.org) or Copernicus (https://www.copernicus.eu/).

5.2 Coordinate the comparison of data sources to assess their uncertainties and value for hydrological research

The global datasets mentioned above would complement national information that LSH dataset developers have access to. As these global datasets are recent, their reliability and accuracy for different regions of the world is not well characterised yet (Addor et al., 2018). Using global datasets alongside better-established regional or national datasets would help to assess their value and limitations for hydrological research and applications. Similarly, the comparison of different data products (e.g., remotely sensed products) using a common cloud-based framework highlight their differences and uncertainties. Finally, in addition to assessing uncertainties using the spread among products, several products now provide uncertainty estimates for their own data (e.g., Newman et al., 2015b; Cornes et al., 2018; Hengl et al., 2017; Chaney et al., 2019), and recent coordinated efforts provide guidance on how to conduct streamflow uncertainty assessments in diverse environments (e.g., Kiang et al., 2018). Together, these methods will enable us to better characterise uncertainties in LSH datasets.

We think there is a need for a platform centralising hydrological information relevant for LSH, on which the value of these global datasets is discussed. It would provide an overview of existing products, a description of the regions covered, the data included, and a discussion of their strengths and limitations (e.g., following the example of the Climate Data Guide, https://climatedataguide.ucar.edu/).

5.3 Sustain efforts to characterise human impacts on streamflow

The level of detail and diversity of geophysical datasets is increasing rapidly, but the characterisation of human impacts is progressing much slower. Although it is difficult to access reliable and consistent water use data, there is an opportunity to use recently released global datasets, such as the Global Reservoir and Dam (GRanD) database (Lehner et al., 2011), global gridded water withdrawals (Huang et al, 2018) and land cover datasets to incorporate and classify the various types of water engineering infrastructure and human-induced land changes in LSH datasets. These attributes would help for an improved understanding of human impacts on hydrological catchment functioning over time.

Several authors have stressed the need to develop indices linking water resources and society (Wada et al., 2017), including a threshold value to characterise the degree of water scarcity (Falkenmark, 1989), the Water Exploitation Index (De Roo et al, 2012), the Blue Water Sustainability Indicator (Wada and Bierkens, 2014) and a human intervention degree index (Alvarez-Garreton et al, 2018). We argue that the inclusion of these and new human intervention indices should be established and standardised, such that meaningful comparisons of human alteration effects can be achieved across catchments globally.

5.4 Increase the accessibility of streamflow data

The accessibility of data varies strongly geographically. A concerted effort is needed by hydrologists and umbrella organisations (e.g., GRDC, IAHS, WMO) to lobby for the public release of currently inaccessible streamflow datasets, particularly in regions where there is little streamflow data readily available (e.g., south-eastern Asia, central Africa). Technological issues mean that historic data may only be available in hard copy or an outdated format, and resources may be unavailable locally to transcribe or convert it. Thus, financial assistance as part of international collaborations could catalyse data sharing. Also, the WMO has prepared a guide for the rescue of such data, and interested agencies are directed to WMO (2014) for guidance on good practice (see also Brönnimann et al., 2018). Clear articulation of local benefits must also be outlined. For example, releasing data for inclusion in large-sample datasets ensures that the geographic region is examined by every study adopting the dataset, yielding operationally-significant insights into the regional hydrology at little cost to the nation or agency. This may partially offset the perception that releasing the data means the loss of a strategic asset.

Overall, there is a need to increase the accessibility and comparability of both observed and simulated streamflow time series. The website http://camels.cr2.cl (Alvarez-Garreton et al., 2018) provides an example of LSH data provision with a high degree of user interaction. We advocate for more hydrological simulations to be shared, in order to facilitate model comparison, benchmarking and improvement (e.g., Best et al., 2015; Newman et al., 2017). The platform Catch X (https://ewgis.org/catchx-global/) provides simulated runoff across 57,646 catchments using global-scale simulations available through the eartH2Observe project (Schellekens et al., 2017). This recently launched platform also includes other hydrometeorological variables (e.g., evapotranspiration, snowfall, temperature) and land cover, and could potentially be one of the toolsets to further bridge the gap between LSH and large-scale hydrology.

6. Conclusions

LSH datasets have enabled progress in multiple fields of hydrological sciences and they are supporting the emergence of novel approaches to better understand water dynamics, relying for instance on machine-learning (Section 1). The content and spatial extent of LSH datasets has significantly expanded over the last decade, and the overview provided in Section 2 and Table 2 should help users to select the datasets corresponding best to their needs. Overall, as new mechanisms are implemented to acknowledge datasets in published studies, the recognition of the key role played by datasets in scientific advances is improving.

Yet, we argue here that to sustain the contribution of LSH datasets to hydrological sciences and to widen the scope of LSH studies, it is essential to better coordinate the production of LSH datasets worldwide (Figure 2). Currently, their use and interpretation is hindered by their lack of comparability, uncertainty estimates and characterisation of human impacts, as well as by the still limited access to streamflow data (Section 3). To overcome these limitations, we propose a list of simple actions that can be taken today when producing or updating a LSH dataset (Section 4). Following these guidelines will increase the overall value of LSH datasets for the community. We argue that to truly overcome the challenges LSH is facing, there is additionally a need for community-wide, longer-term efforts (Section 5). We propose in particular to move the production of the LSH datasets to the cloud, in order to accelerate their standardisation and facilitate their future management.

Following the guidelines and focussing on the grand challenges we outline in this paper has the potential to enhance the transparency and reproducibility of hydrological studies, and to lead better structured, less fragmented LSH datasets. We believe that hydrologists will benefit from using data from a common, large sample of catchments, in order to test hydrological hypotheses and models. Relying on and contributing to such a shared resource will significantly increase the comparability of individual studies, and thereby, enhance our ability to learn from their combined results.

Acknowledgements

This study is a contribution to the Large-sample Hydrology working group of the Panta Rhei Research Initiative of the International Association of Hydrological Sciences (IAHS). We thank Lukas Gudmundsson and the participants of the splinter meeting “Large-sample hydrology: facilitating the production and exchange of datasets worldwide” at EGU2018 for their inputs. NA is supported by the Swiss National Science Foundation (Fellowships P2ZHP2_161963 and P400P2_180791). HXD acknowledges supports from the Australia Awards Scholarship (ST000HKE6) and the University of Adelaide D R Stranks Fellowship. CAG is funded by FONDECYT postdoctoral grant no. 3170428 and the Center for Climate and Resilience Research (CR2, CONICYT/FONDAP/15110009). GC is supported by NERC MaRIUS: Managing the Risks, Impacts and Uncertainties of droughts and water Scarcity (grant NE/L010399/1). KF acknowledges the support of the Bureau of Meteorology, Australia (TP705654) and the Australian Research Council (LP170100598). PAM acknowledges support from FONDECYT postdoctoral grant No. 3170079.

References

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21(10), 5293–5313, doi:10.5194/hess-21-5293-2017, 2017.

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N. and Clark, M. P.: A ranking of hydrological signatures based on their predictability in space, Water Resour. Res., doi:10.1029/2018WR022606, 2018.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Galleguillos, M., Zambrano, M., Lara, A., Addor, N., Puelma, C., Cortes, G., Garreaud, R., McPhee, J. and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies - Chile dataset, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2018-23, 2018.

Andréassian, V., Hall, A., Chahinian, N. and Schaake, J.: Introduction and Synthesis : Why should hydrologists work on a large number of basin data sets ?, Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment–MOPEX. IAHS Publ. 307, 1–5, 2006a.

Andreassian, V., Hall, A., Chahinian, N., & Schaake, J. (Eds): Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment–MOPEX. IAHS Publ. 307, pp 348 + iv, 2006b.

Arsenault, R., Bazile, R., Ouellet Dallaire, C., and Brissette, F.: CANOPEX: A Canadian hydrometeorological watershed database, Hydrological Processes, 30, 2734-2736, 2016.

Australian National Data Service (ANDS), Research Data Services (RDS) and the National eResearch Collaboration Tools and Resources Project (NECTAR): FAIR self assessment tool, 2017. Available online at https://www.ands-nectar-rds.org.au/fair-tool

Barbarossa, V., Huijbregts, M. A. J., Beusen, A. H. W., Beck, H. E., King, H. and Schipper, A. M.: FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015, Sci. data, 5, doi:10.1038/sdata.2018.78, 2018.

Beck, H. E., de Roo, A., and van Dijk, A. I. J. M.: Global maps of streamflow characteristics based on observations from several thousand catchments, Journal of Hydrometeorology, doi: 10.1175/JHM-D-14-0155.1, 2015.

Beck, H. E., van Dijk, A. I. J. M., Roo, A. de, Miralles, D. G., McVicar, T. R., Schellekens, J. and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52(5), 3599–3622, doi:10.1002/2015WR018247, 2016.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G., Brocca, L., Pappenberger, F., Huffman, G. J. and Wood, E. F.: Global-scale evaluation of 23 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth Syst. Sci., 21, 6201–6217, doi:10.5194/hess-2017-508, 2017.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. M., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment, Bulletin of the American Meteorological Society, in press, 2019.

Bennett, B., Leonard, M., Deng, Y., and Westra, S.: An empirical investigation into the effect of antecedent precipitation on flood volume, Journal of Hydrology, doi: https://doi.org/10.1016/j.jhydrol.2018.10.025, 2018.

Berghuijs, W. R., Aalbers, E. E., Larsen, J. R., Trancoso, R. and Woods, R. A.: Recent changes in extreme floods across multiple continents, Environ. Res. Lett., 12(11), 22–25, doi:10.1088/1748-9326/aa8847, 2017.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. ., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. A., Stevens, L. and Vuichard, N.: The plumbing of land surface models: benchmarking model performance, J. Hydrometeorol., 16, 1425–1442, doi:10.1175/JHM-D-14-0158.1, 2015.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H. and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrol. Process., 29, 310–320, doi:10.1002/hyp.10391, 2015.

Bierkens, M. F. P.: Global hydrology: State, trends, and directions, Water Resour. Res., 51(7), 4923–4947, doi:10.1002/2015WR017173, 2015.

Blenkinsop S, Lewis E, Chan SC, Fowler HJ, Quality control of an hourly rainfall dataset and climatology of extremes for the UK. Int J Climatol 37(2):722–740. https://doi.org/10.1002/joc.4735, 2017.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Ivan, Č., Castellarin, A. and Chirico, G. B.: Changing climate shifts timing of European floods, Science (80-.)., 357(6351), 588–590, doi:10.1126/science.aan2506, 2017.

Brönnimann, S., Brugnara, Y., Allan, R. J., Brunet, M., Compo, G. P., Crouthamel, R. I., Jones, P. D., Jourdain, S., Luterbacher, J., Siegmund, P., Valente, M. A. and Wilkinson, C. W.: A roadmap to climate data rescue services, Geosci. Data J., 5(1), 28–39, doi:10.1002/gdj3.56, 2018.

Brown, A. E., Zhang, L., McMahon, T. A., Western, A. W., & Vertessy, R. A.: A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation, Journal of hydrology, 310(1-4), 28-61, 2005.

Burn, D. H. and Whitfield, P. H.: Changes in flood events inferred from centennial length streamflow data records, Advances in Water Resources, 121, 333-349, 2018.

Ceola, S., Arheimer, B., Baratti, E., Blöschl, G., Capell, R., Castellarin, A., Freer, J., Han, D., Hrachowitz, M., Hundecha, Y., Hutton, C., Lindström, G., Montanari, A., Nijzink, R., Parajka, J., Toth, E., Viglione, A. and Wagener, T.: Virtual laboratories: New opportunities for collaborative water science, Hydrol. Earth Syst. Sci., 19(4), 2101–2117, doi:10.5194/hess-19-2101-2015, 2015.

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C., Morgan, C. L. S., McBratney, A. B., Wood, E. F. and Yimam, Y. T.: POLARIS soil properties: 30-meter probabilistic maps of soil properties over the contiguous United States, Water Resour. Res., 2018WR022797, doi:10.1029/2018WR022797, 2019.

Cloke, H.L. and Hannah, D.M.. Large-scale hydrology: advances in understanding processes, dynamics and models from beyond river basin to global scale. Hydrological Processes, 25(7), pp.991-995, 2011.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M. and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, J. Geophys. Res. Atmos., 123(17), 9391–9409, doi:10.1029/2017JD028200, 2018.

Coron L, Andréassian V, Perrin C, Lerat J, Vaze J, Bourqui M, Hendrickx F. Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. Water Resources Research 48, W05552 DOI:10.1029/2011WR011721, 2012.

Coxon, G., Freer, J., Westerberg, I.K., Wagener, T., Woods, R. and Smith, P.J., A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. Water Resources Research 51(7), 5531-5546, 2015.

Coxon, G., Freer, J., Wagener, T., Odoni, N. A. and Clark, M.: Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrol. Process., 28(25), 6135–6150, doi:10.1002/hyp.10096, 2013.

De Roo, A., Burek, P., Gentile, A., Udias, A., Bouraoui, F., Aloe, A., Bianchi, A., La Notte, A., Kuik, O., Tenreiro, J. E., Vandecasteele, I., Mubareka, S., Baranzelli, C., Der Perk, M. V., Lavelle, C., and Bidoglio, G.: A multi-criteria optimisation of scenarios for the protection of water resources in Europe, JRC Scientific and Policy Report, JRC75919, ISSN: 1831-9424, EC-JRC-IES, Italy, 2012.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765-785, 2018a.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive - Part 1: Station catalog and Catchment boundary. PANGAEA, https://doi.org/10.1594/PANGAEA.887477, 2018b.

Do, H. X., Westra, S., and Michael, L.: A global-scale investigation of trends in annual maximum streamflow, Journal of Hydrology, doi: 10.1016/j.jhydrol.2017.06.015, 2017.

Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H.V., Gusev, Y.M., Habets, F., Hall, A., Hay, L. and Hogue, T.,: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. Journal of Hydrology, 320(1-2), pp.3-17, 2006.

Falkenmark, M.: The massive water scarcity now threatening Africa-why isn't it being addressed?, Ambio, 18, 112–118, 1989.

Falkenmark, M. and Chapman, T. (Eds.): Comparative Hydrology, UNESCO, Paris, 1989.

Fekete, B. M., Vörösmarty, C. J. and Wolfgang, G.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cycles, 16(3), doi:10.1029/1999GB001254, 2002.

Ferguson, I. M., & Maxwell, R. M.: Human impacts on terrestrial hydrology: Climate change versus pumping and irrigation. Environmental Research Letters, 7(4), 044022. https://doi.org/10.1088/1748-9326/7/4/044022, 2012.

Finlayson B., McMahon T., Srikanthan R. & Haines A.: World Hydrology: a new data base for comparative analysis, Proceedings of the Hydrology and Water Resources Symposium, Nat. Conf. Publ. 86/13, The Institution of Engineers, pp. 288-296, 1986.

Fowler, K. Peel, M., Western, A., Zhang, L. and Peterson, T.: Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models, Water Resources Research, 52, 1820–1846, doi:10.1002/2015WR018068, 2016.

Fowler, K., Peel, M., Western, A., and Zhang, L.: Improved Rainfall-Runoff Calibration for Drying Climate: Choice of Objective Function, Water Resources Research, 54, 3392-3408, 10.1029/2017WR022466, 2018.

Ghiggi, G., Humphrey, V., Seneviratne, S. I. and Gudmundsson, L.: GRUN: An observations-based global gridded runoff dataset from 1902 to 2014, Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2019-32, 2019.

Gleeson, T., Moosdorf, N., Hartmann, J. and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41, 3891–3898, doi:10.1002/2014GL059856, 2014.

Grabs, W.: Information transfer in hydrology: experiences of the Global Runoff Data Centre. In: FRIEND '97 — Regional Hydrology: Concepts and Models for Sustainable Water Resource Management (Proceedings of the Postojna, Slovenia, Conference, September–October 1997). IAHS Publ. No. 246, pp. 13–19, 1997

GRDC: Report of the Twelfth Meeting of the GRDC Steering Committee, Koblenz, Germany, 18–19 June 2014, Global Runoff Data Centre (GRDC), Koblenz, Germany, 23 pp., 2015.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F. and Koirala, S.: Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, J. Hydrometeorol., 13(2), 604–620, doi:10.1175/JHM-D-11-083.1, 2012.

Gudmundsson, L. and Seneviratne, S. I.: Observation-based gridded runoff estimates for Europe (E-RUN version 1.1), Earth Syst. Sci. Data, 8(2), 279–295, doi:10.5194/essd-8-279-2016, 2016.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787-804, 2018a.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) - Part 2: Time Series Indices and Homogeneity Assessment. PANGAEA, https://doi.org/10.1594/PANGAEA.887470, 2018b

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., & Seneviratne, S. I. (2019). Observed trends in global indicators of mean and extreme streamflow. Geophysical Research Letters, 46, 756–766. https://doi.org/10.1029/2018GL079725

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18, 463-477, https://doi.org/10.5194/hess-18-463-2014, 2014.

Hanasaki, N., S. Kanae, & T. Oki: A reservoir operation scheme for global river routing models. J. Hydrol., 327, 22–41, 2006.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, Hydrological Processes, 25, 1191-1200, 2011.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K., and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrol. Earth Syst. Sci., 22, 2023-2039, https://doi.org/10.5194/hess-22-2023-2018, 2018.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophys. Geosystems, 13(12), 1–37, doi:10.1029/2012GC004370, 2012.

Hengl, T., Mendes De Jesus, J., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotí, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. and Kempen, B.: SoilGrids250m: Global Gridded Soil Information Based on Machine Learning, PLoS One, 12(2), e0169748, doi:10.1371/journal.pone.0169748, 2017.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E. and Dadson, S. J.: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, J. Hydrol., 566, 595–606, doi:10.1016/j.jhydrol.2018.09.052, 2018.

Horsburgh, J. S., Tarboton, D. G., Piasecki, M., Maidment, D. R., Zaslavsky, I., Valentine, D. and Whitenack, T.: An integrated system for publishing environmental observations data, Environ. Model. Softw., 24(8), 879–888, doi:10.1016/j.envsoft.2009.01.002, 2009.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. a., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. a., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. a., Zehe, E. and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrol. Sci. J., 58(6), 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C. and Arheimer, B.: Most computational hydrology is not reproducible, so is it really science?, Water Resour. Res., 52, 7548–7555, doi:10.1002/2016WR019285, 2016.

Jaramillo, F., & Destouni, G.: Local flow regulation and irrigation raise global human water consumption and footprint. Science, 350(6265), 1248–1251. https://doi.org/10.1126/science.aad1010, 2015.

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P. and Thielen, J.: Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level, Environ. Model. Softw., 75, 68–76, doi:10.1016/j.envsoft.2015.09.009, 2016.

Keller, V. D. J., Tanguy, M., Prosdocimi, I., Terry, J. A., Hitt, O., Cole, S. J., Fry, M., Morris, D. G., and Dixon, H.: CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications, Earth Syst. Sci. Data, 7, 143-155, https://doi.org/10.5194/essd-7-143-2015, 2015.

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., et al. A comparison of methods for streamflow uncertainty estimation. Water Resources Research, 54, 7149–7176. https://doi.org/10.1029/2018WR022708, 2018.

Kirchner, J.W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42, n/a–n/a. doi:10.1029/2005WR004362, 2006.

Kovács, G.: Proposal to construct a coordinating matrix for comparative hydrology, Hydrol. Sci. J., 29(4), 435–443, doi:10.1080/02626668409490961, 1984.

Knoben, W. J. M., Woods, R. A. and Freer, J. E.: A Quantitative Hydrological Climate Classification Evaluated with Independent Streamflow Data, Water Resour. Res., 54, doi:10.1029/2018WR022913, 2018.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger, M.: Rainfall-Runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., doi:10.5194/hess-22-6005-2018, 2018.

Kuentz, A., Arheimer, B., Hundecha, Y. and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrol. Earth Syst. Sci., 21, 2863–2879, doi:10.5194/hess-21-2863-2017, 2017.

Lehner, B., C. Reidy Liermann, C. Revenga, C. Vorosmarty, B. Fekete, P. Crouzet, P. Doll, M. Endejan, K. Frenken, J. Magome, C. Nilsson, J.C. Robertson, R. Rodel, N. Sindorf, and D. Wisser: Global Reservoir and Dam Database, Version 1 (GRanDv1): Dams, Revision 01. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), https://doi.org/10.7927/H4N877QK, 2011.

Lehner, B.. Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network. Global Runoff Data Centre in the Federal Institute of Hydrology (BFG), 2012.

Lewis, Elizabeth, et al., A Rule Based Quality Control Method for Hourly Rainfall Data and a 1 km Resolution Gridded Hourly Rainfall Dataset for Great Britain: CEH-GEAR1hr, Journal of hydrology, 564, pp. 930-943. doi: 10.1016/j.jhydrol.2018.07.034, 2018.

Liu, X., Tang, Q., Cui, H., Mu, M., Gerten, D., Gosling, S., Masaki, Y., Wada, Y. and Satoh, Y.: Multimodel uncertainty changes in simulated river flows induced by human impact parameterizations, Environ. Res. Lett. 12 025009, 2017.

Mallakpour, I. and Villarini, G., (2015). The changing nature of flooding across the central United States. Nature Climate Change, 5(3), p.250, 2015.

McMahon T.: Hydrological Characteristics of Selected Rivers of the World, Technical documents in hydrology, Unesco, Paris, 1982.

McMillan, H., Krueger, T. and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrological Processes 26(26), 4078-4111, 2012.

McMillan, H. K., Booker, D. J. and Cattoën, C.: Validation of a national hydrological model, J. Hydrol., 541, 800–815, doi:10.1016/j.jhydrol.2016.07.043, 2016a.

McMillan, H., Montanari, A., Cudennec, C., Savenije, H., Kreibich, H., Krueger, T., Liu, J., Mejia, A., Van Loon, A., Aksoy, H., Di Baldassarre, G., Huang, Y., Mazvimavi, D., Rogger, M., Sivakumar, B., Bibikova, T., Castellarin, A., Chen, Y., Finger, D., Gelfan, A., Hannah, D. M., Hoekstra, A. Y., Li, H., Maskey, S., Mathevet, T., Mijic, A.,

Acuña, A. P., Polo, M. J., Rosales, V., Smith, P., Viglione, A., Srinivasan, V., Toth, E., van Nooyen, R. and Xia, J.: Panta Rhei 2013-2015: Global perspectives on hydrology, society and change, Hydrol. Sci. J., 61(7), 1174–1191, doi:10.1080/02626667.2016.1159308, 2016b.

Melsen, L. A., Torfs, P. J. J. ., Uijlenhoet, R. and Teuling, A. J.: Comment on "Most computational hydrology is not reproducible, so is it really science?," Water Resour. Res., 52(10), 7548–7555, doi:10.1002/2016WR019285, 2016.

Melsen, L., Addor, N., Mizukami, N., Newman, A., Torfs, P., Clark, M., Uijlenhoet, R. and Teuling, R.: Mapping (dis)agreement in hydrologic projections, Hydrol. Earth Syst. Sci., 22, 1775–1791, doi:10.5194/hess-22-1775-2018, 2018.

Moine, M. P., Valcke, S., Lawrence, B. N., Pascoe, C., Ford, R. W., Alias, A., Balaji, V., Bentley, P., Devine, G., Callaghan, S. A. and Guilyardi, E.: Development and exploitation of a controlled vocabulary in support of climate modelling, Geosci. Model Dev., 7(2), 479–493, doi:10.5194/gmd-7-479-2014, 2014.

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S. J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D. A., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z. and Belyaev, V.: "Panta Rhei-Everything Flows": Change in hydrology and society-The IAHS Scientific Decade 2013-2022, Hydrol. Sci. J., 58(6), 1256–1275, doi:10.1080/02626667.2013.809088, 2013.

National Research Council (NRC): Bits of Power: Issues in Global Access to Scientific Data, National Academy Press, Washington, DC, 1997

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T. and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous USA: dataset characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, doi:10.5194/hess-19-209-2015, 2015a.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L. and Arnold, J. R.: Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States, J. Hydrometeorol., 16, 2481–2500, doi:10.1175/JHM-D-15-0026.1, 2015b.

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B. and Nearing, G.: Benchmarking of a physically based hydrologic model, J. Hydrometeorol., JHM-D-16-0284.1, doi:10.1175/JHM-D-16-0284.1, 2017.

Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., van Nooijen, R. R. P., Savenije, H. H. G. and Hrachowitz, M.: Constraining conceptual hydrological models with multiple information sources, Water Resour. Res., 54, doi:10.1029/2017WR021895, 2018.

Open Data Charter: International Open Data Charter, 2015. Available online at https://opendatacharter.net/wp-content/uploads/2015/10/opendatacharter-charter_F.pdf

Oudin, L., Kay, A., Andréassian, V. and Perrin, C.: Are seemingly physically similar catchments truly hydrologically similar?, Water Resour. Res., 46(W11558), doi:10.1029/2009WR008887, 2010.

Peel, M., McMahon, T., Finlayson, B., Watson, F.: Identification and explanation of continental differences in the variability of annual runoff. Journal of Hydrology, 250, 224-240, 2001.

Peel M., McMahon T. & Finlayson B.: Continental differences in the variability of annual runoff - update and reassessment. Journal of Hydrology. 295, 185-197, 2004.

Pelletier, J. D., Patrick D. Broxton, Hazenberg, P., Zeng, X., Troch, P. A., Niu, G.-Y., Williams, Z., Brunke, M. A. and Gochis, D.: A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling, J. Adv. Model. Earth Syst., 8, doi:10.1002/2015MS000526, 2016.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M. and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, J. Hydrometeorol., 17(1), 287–307, doi:10.1175/JHM-D-15-0054.1, 2016.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the
eastern USA, Hydrol. Earth Syst. Sci., 15, 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B.,

Burke, S., Dorigo, W. and Weedon, G. P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth Syst. Sci. Data, 9, 389–413, doi:10.5194/essd-2016-55, 2017.

Seibert, J., Vis, M., Lewis, E. and van Meerveld, I.: Upper and lower benchmarks in hydrological modeling, Hydrol. Process., 32, 1120–1125, doi:10.1002/hyp.11476, 2018.

Singh, R., Archfield, S. A. and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments - A comparative hydrology approach, J. Hydrol., 517, 985–996, doi:10.1016/j.jhydrol.2014.06.030, 2014.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J.W., Schertzer, D., Uhlenbrook, S., Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrol. Sci. J. 48, 857–880. doi:10.1623/hysj.48.6.857.51421, 2003.

Sivapalan, M.: The secret to 'doing better hydrological science': change the question!, Hydrol. P, 23, 1391–1396, doi:10.1002/hyp, 2009.

Slater, L. J., Thirel, G., Harrigan, S., Delaigue, O., Hurley, A., Khouakhi, A., Prodoscimi, I., Vitolo, C. and Smith, K.: Using R in hydrology: a review of recent developments and future directions, Hydrol. Earth Syst. Sci. Discuss, doi:10.5194/hess-2019-50, 2019.

Slater, L. J., & Villarini, G.: Enhancing the predictability of seasonal streamflow with a statistical-dynamical approach, Geophys. Res. Lett., 45(13), 6504-6513, doi:10.1029/2018GL077945, 2018.

Stoelzle, M., Stahl, K. and Weiler, M.: Are streamflow recession characteristics really characteristic?, Hydrol. Earth Syst. Sci., 17(2), 817–828, doi:10.5194/hess-17-817-2013, 2013.

Thompson, S. E., Harman, C. J., Konings, A. G., Sivapalan, M., Neal, A. and Troch, P. A.: Comparative hydrology across AmeriFlux sites: The variable roles of climate, vegetation, and groundwater, Water Resour. Res., 47(7), W00J07, doi:10.1029/2010WR009797, 2011.

Tijdeman, E., Bachmair, S. and Stahl, K.: Controls on hydrologic drought duration in near-natural streamflow in Europe and the USA, Hydrol. Earth Syst. Sci., 20(10), 4043–4059, doi:10.5194/hess-20-4043-2016, 2016.

Tijdeman, E., Hannaford, J. and Stahl, K.: Human influences on streamflow drought characteristics in England and Wales, Hydrol. Earth Syst. Sci., 22(2), 1051–1064, doi:10.5194/hess-22-1051-2018, 2018a.

Tijdeman, E., Barker, L. J., Svoboda, M. D. and Stahl, K.: Natural and Human Influences on the Link Between Meteorological and Hydrological Drought Indices for a Large Set of Catchments in the Contiguous United States, Water Resour. Res., 54(9), 6005–6023, doi:10.1029/2017WR022412, 2018b.

Troch, P. A., Martinez, G. F., Pauwels, V. R. N., Durcik, M., Sivapalan, M., Harman, C., Brooks, P. D., Gupta, H. and Huxman, T.: Climate and vegetation water use efficiency at catchment scales, Hydrol. Process., 23, 2409–2414, doi:10.1002/hyp.7358, 2009.

van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J., and Beck, H. E.: Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide, Water Resources Research, 49, 2729-2746, 2013.

Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J. & Satoh, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. Environmental Research Letters, *13*(5), 055008, 2018.

Vertessy, R. A.: Impacts of plantation forestry on catchment runoff, in E. K. Sadanandan Nabia and A. G. Brown (Eds.), Plantations, Farm Forestry and Water, Proceedings of a National Workshop, 20–21 July, Melbourne, pp. 9–19, 2000.

Viglione, A., Borga, M., Balabanis, P. and Bloschl, G.: Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy. J. Hydrol. 394(1-2), 63-77, doi:10.1016/j.jhydrol.2010.03.023, 2010.

Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C. J. A. and Buytaert, W.: Web technologies for environmental Big Data, Environ. Model. Softw., 63, 185–198, doi:10.1016/j.envsoft.2014.10.007, 2015.

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B.: Global water resources: Vulnerability from climate change and population growth. Science, 289, 284–288, 2000.

Wada, Y., & Bierkens, M. F.: Sustainability of global water use: past reconstruction and future projections. Environmental Research Letters, 9(10), 104003, 2014.

Wada, Y., Bierkens, M. F. P., de Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., Konar, M., Liu, J., Müller Schmied, H., Oki, T., Pokhrel, Y., Sivapalan, M., Troy, T. J., van Dijk, A. I. J. M., van Emmerik, T., Van Huijgevoort, M. H. J., Van Lanen, H. A. J., Vörösmarty, C. J., Wanders, N., and Wheater, H.: Human–water interface in hydrological modelling: current status and future directions, Hydrol. Earth Syst. Sci., 21, 4169-4193, https://doi.org/10.5194/hess-21-4169-2017, 2017.

Wehren, B., B. Schädler, and R. Weingartner: Human interventions, in U. Bundi, et al. (Eds.), Alpine Waters, The Handbook of Environmental Chemistry, Volume 6, Heidelberg: Springer Verlag, pp. 71–92, 2010.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resources Research, 52, 1847–1865, https://doi.org/10.1002/2015WR017635, 2016.

Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18, 2016.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A. and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47(5), W05301, doi:10.1029/2010WR010090, 2011.

World Meteorological Organization (WMO): International Glossary of Hydrology. WMO-No. 385, (2012). ISBN 978-92-63-03385-8, Geneva 2, Switzerland, available online: https://library.wmo.int/pmb_ged/wmo_385-2012.pdf, 2012.

World Meteorological Organization (WMO): Guidelines for Hydrological Data Rescue. WMO-No. 1146, (2014). ISBN 978-92-63-11146-3, Geneva 2, Switzerland, available online: http://www.wmo.int/pages/prog/hwrp/publications/guidelines_hydrological_DR/wmo_1146_en.pdf, 2014.

Zaherpour, J., Gosling, S.N., Mount, N., Schmied, H.M., Veldkamp, T.I., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I. and Hanasaki, N., (2018). Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. Environmental Research Letters, 13(6), p.065015, 2018.

Zink, M., Kumar, R., Cuntz, M., and Samaniego, L.: A high-resolution dataset of water fluxes and states for Germany accounting for parametric uncertainty, Hydrol. Earth Syst. Sci., 21, 1769-1790, https://doi.org/10.5194/hess-21-1769-2017, 2017.

Zhang, L., Dobslaw, H., Stacke, T., Güntner, A., Dill, R., and Thomas, M.: Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations, Hydrol. Earth Syst. Sci., 21, 821-837, https://doi.org/10.5194/hess-21-821-2017, 2017.

Table 1 Summary of the range of data potentially included in LSH datasets. Italic text highlights data currently underrepresented.

	Minimum requirement	**Additional information**	**Metadata and uncertainty estimates**
Streamflow observations and basic catchment identifiers	Daily or monthly streamflow records Basic identifiers (name, gauge coordinates, catchment area)	Hydrological signatures GIS layer of catchment boundary	Quality flags for observations *Flag if station included in other LSH or "reference" (high quality) datasets* *Rating curve information* *Infos on gauge relocation* *Uncertainty estimates*
Hydrometeorological time series	No minimum requirements	*Precipitation* *Temperature* *Potential Evapotranspiration* *Evapotranspiration* *Soil moisture* *Snow water equivalent*	*Discussion of data sources limitations* *Uncertainty estimates*
Landscape and hydroclimatic attributes	No minimum requirements	Topography Climate Land cover Soil Geology *Stream network* *Human impacts (abstractions, irrigation, dams, human population)*	References to nested or neighbour catchments *Discussion of data sources limitations* *Uncertainty estimates*

Table 2 Overview of key LSH datasets currently available. *Streamflow (Q), precipitation (P), temperature (T), potential evapotranspiration (PET) and snow water equivalent (SWE)

1161

Name	Number of catchments	Location	Hydro-meteorological time series*	Basic catchment identifiers	Catchment attributes	Availability	DOI or URL for data set download	Reference or provider
The Global Streamflow Indices and Metadata Archive (GSIM)	35,002	Global	Monthly, seasonal and annual streamflow indices	Name, gauge coordinates, catchment area and boundary	Topography, land cover, geology, irrigation, human population, soil profile	Freely available under Creative Commons Attribution-ShareAlike 3.0 Unported license	https://doi.org/10.1594/PANGAEA.887477 https://doi.org/10.1594/PANGAEA.887470	Do et al. (2018) Gudmundsson et al. (2018)
The Global Runoff Data Base (GRDB)	9543	Global	Q	Name, gauge coordinates, catchment area. Catchment boundary available for some stations	No catchment attributes available	Free and unrestricted (but identified) access. No redistribution of the data by the user.	https://www.bafg.de/GRDC/	GRDC
European Water Archive (EWA)	3731	Europe	Q	Name, gauge coordinates, catchment area. Catchment boundary available for some stations	No catchment attributes available	Free and unrestricted (but identified) access. No redistribution of the data by the user.	http://ne-friend.bafg.de/servlet/is/7398/ http://www.bafg.de/GRDC/EN/04_spcldtbss/42_EWA/ewa_node.html	GRDC
European catchments of Hydrological Predictions for the Environment model (E-HYPE)	35215 catchments 1366 river gauges	Europe	Q	Name, gauge coordinates, catchment area and boundary	Topography, land cover, geology, soil profile	Freely available under Creative Commons Attribution Share Alike 4.0 International license	https://doi.org/10.5281/zenodo.581428 https://doi.org/10.5281/zenodo.581429 https://doi.org/10.5281/zenodo.581430 https://doi.org/10.5281/zenodo.581431 https://doi.org/10.5281/zenodo.581432 https://doi.org/10.5281/zenodo.581434 https://doi.org/10.5281/zenodo.581435	Kuentz et al (2017)
Model Parameter Estimation Experiment (MOPEX)	438	US	P, T, Q	Name, gauge coordinates, catchment area and boundary	Topography, climatic indices, land cover, soil, geology	Freely available	http://www.nws.noaa.gov/oh/mopex/	Schaake et al. (2006)
Canadian model parameter experiment (CANOPEX)	698	Canada	P, T, Q	Name, gauge coordinates, catchment area and boundary	No catchment attribute available	Freely available	http://canopex.etsmtl.net	Arsenault et al. (2016)
Catchment Attributes and MEteorology for Large-sample Studies dataset (CAMELS)	671	US	P, T, Q, PET	Name, gauge coordinates, catchment area and boundary	Topography, climatic indices, hydrological signatures, land cover, soil, geology	Freely available	https://dx.doi.org/10.5065/D6MW2F4D https://dx.doi.org/10.5065/D6G73C3Q	Newman et al. (2015) Addor et al. (2017)
Catchment attributes and meteorology for large-sample studies dataset for Chile (CAMELS-CL)	516	Chile	P, T, Q, PET, SWE	Name, gauge coordinates, catchment area and boundary	Topography, climatic indices, hydrological signatures, land cover, geology, water use	Freely available under Creative Commons Attribution 4.0 International	https://doi.org/10.1594/PANGAEA.894885	Alvarez-Garreton et al. (2018)
Regional hydrometeorological data network for the pan-Arctic Region (ARCTICNET)	139	Russia	Q	Name, gauge coordinates, catchment area	No catchment attributes available	Freely available Compiled into GSIM	http://www.r-arcticnet.sr.unh.edu/v4.0/index.html	The State Hydrological Institute, Russia
China Hydrology Data Project (CHDP)	163	China	Q	Name, gauge coordinates, catchment area	No catchment attributes available	Freely provided by author Compiled into GSIM	http://www2.oberlin.edu/faculty/aschmidt/chdp/index.html	Geology Department, Oberlin College, US
GEWEX Asian monsoon experiment – Tropics project (GAME)	129	Thailand	Q	Name, gauge coordinates, catchment area	No catchment attributes available	Freely available Compiled into GSIM	http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/routine/rid-river/index.html	University of Tokyo, Japan

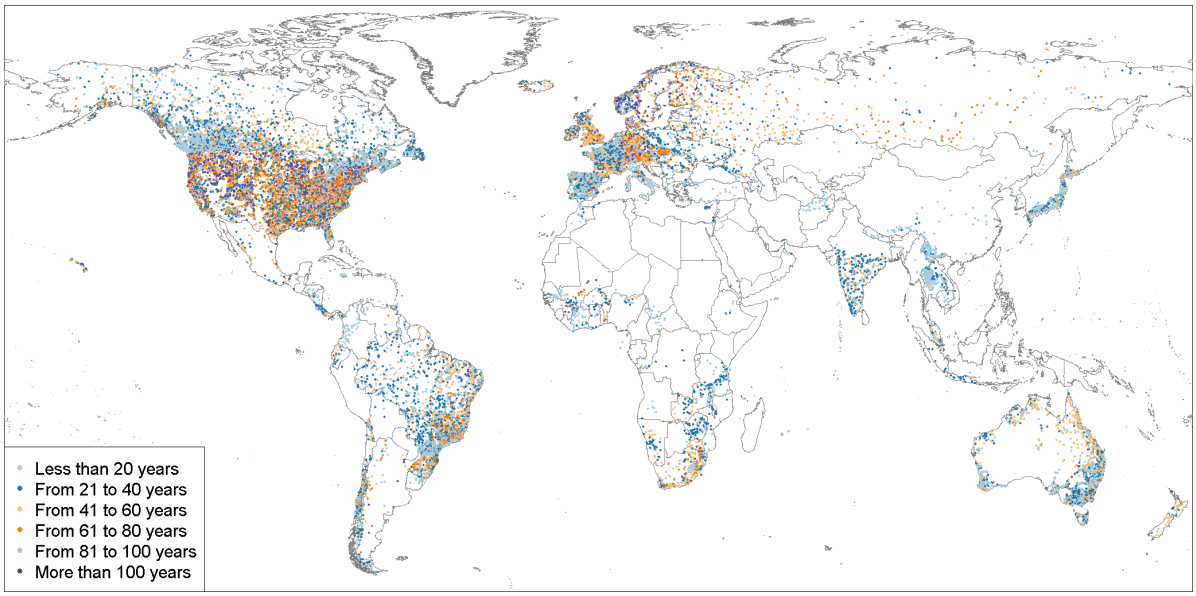


Figure 1 Overview of the location and record length of over 31,500 'findable' stream gauges from the Global Streamflow Indices and Metadata (GSIM) archive and the Catchment Attributes and MEteorology for Large-sample Studies dataset for Chile (CAMELS-CL, which was released after GSIM).

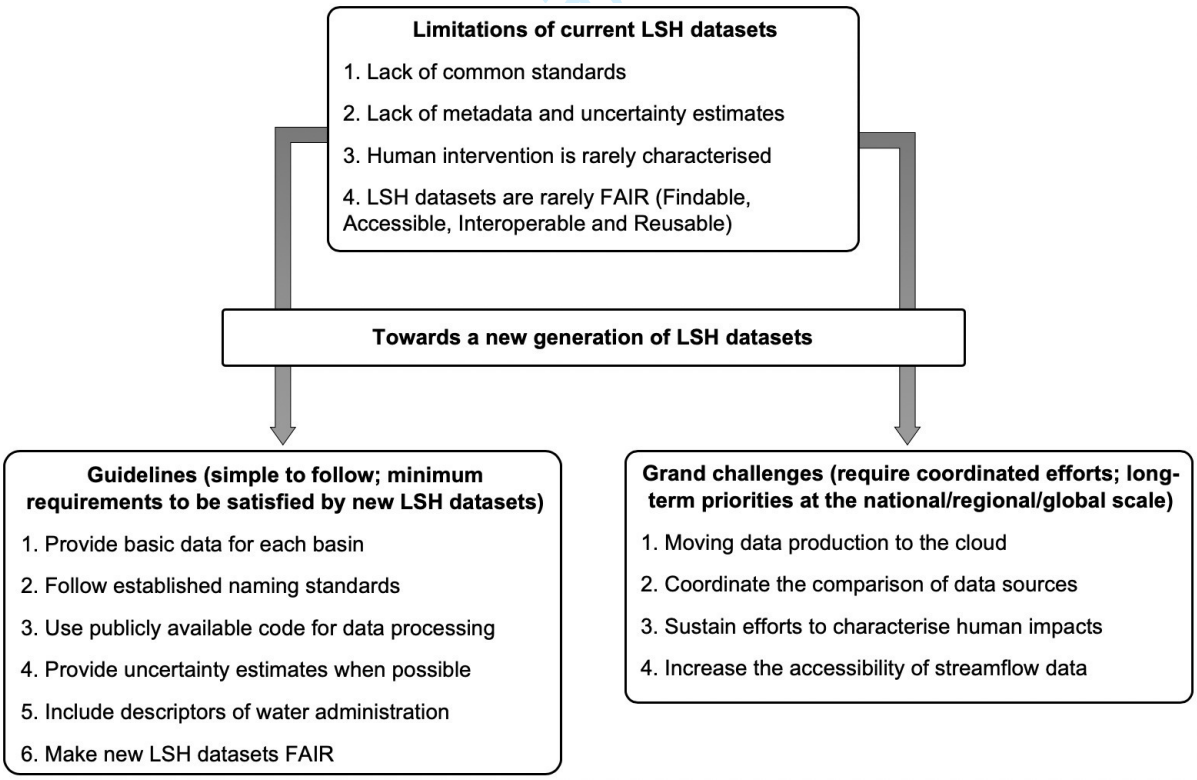


Figure 2 Schematic summary of the limitations of current LSH datasets (Section 3), proposed guidelines for new datasets (Section 4) and grand challenges for LSH (Section 5).

Appendices

Appendix A – Paper 1 (constituting Chapter 2)

Do, H.X., Gudmundsson, L., Leonard, M., and Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of daily streamflow archive and metadata, *Earth Syst. Sci. Data Discuss.*, https://doi.org/10.5194/essd-10-765-2018.

Earth Syst. Sci. Data, 10, 765–785, 2018
https://doi.org/10.5194/essd-10-765-2018

Open Access Earth System Science Data

The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata

Hong Xuan Do[1], Lukas Gudmundsson[2], Michael Leonard[1], and Seth Westra[1]

[1]School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, Australia
[2]ETH Zürich, Institute for Atmospheric and Climate Science, Zürich, Switzerland

Correspondence: Hong Xuan Do (hong.do@adelaide.edu.au)

Received: 7 September 2017 – Discussion started: 20 September 2017
Revised: 9 March 2018 – Accepted: 16 March 2018 – Published: 17 April 2018

Abstract. This is the first part of a two-paper series presenting the Global Streamflow Indices and Metadata archive (GSIM), a worldwide collection of metadata and indices derived from more than 35 000 daily streamflow time series. This paper focuses on the compilation of the daily streamflow time series based on 12 free-to-access streamflow databases (seven national databases and five international collections). It also describes the development of three metadata products (freely available at https://doi.pangaea.de/10.1594/PANGAEA.887477): (1) a GSIM catalogue collating basic metadata associated with each time series, (2) catchment boundaries for the contributing area of each gauge, and (3) catchment metadata extracted from 12 gridded global data products representing essential properties such as land cover type, soil type, and climate and topographic characteristics. The quality of the delineated catchment boundary is also made available and should be consulted in GSIM application. The second paper in the series then explores production and analysis of streamflow indices. Having collated an unprecedented number of stations and associated metadata, GSIM can be used to advance large-scale hydrological research and improve understanding of the global water cycle.

1 Introduction

Streamflow observations with global coverage are essential to make progress in the science of large-scale hydrology. For example, global datasets provide particular value when evaluating global hydrological models (Gudmundsson et al., 2012; Huang et al., 2016; Ward et al., 2013), producing runoff estimation data products (Fekete et al., 2002a, b; Gudmundsson and Seneviratne, 2015; Vörösmarty et al., 1989), investigating large-scale weather patterns and their relation to hydrological extremes (Wanders and Wada, 2015; Ward et al., 2014), and detecting changes in the global hydrological extremes over space and time (Do et al., 2017; Gudmundsson et al., 2017; Kundzewicz et al., 2012; Milly et al., 2002), amongst numerous other applications.

Despite the fundamental, widespread, and varied applications that streamflow observations support, there are many obstacles to the existence and utility of a large-scale streamflow archive. Firstly, there are threats to the quantity of data, such as political sensitivities (Nelson, 2009), cost recovery and strict access policies (Hannah et al., 2011), unavailability in an electronic format, consistency of data formats, limited documentation, missing metadata, and a lack of resources for database maintenance and updating. Secondly, there are difficulties associated with the quality of the data in many regions, such as poor spatial coverage, poor quality control, variable quality control between regions, inconsistent metadata, imprecise geographic coordinates of the site, changes in the density of stream gauges, and variable record lengths. Lastly, even in locations where there are abundant and high-quality streamflow observations, there can be questions over its utility in specific research such as climate sensitivity analysis due to the manifestation of human impacts – for example, urbanization, land-use changes, channelization, and upstream dams (Hannah et al., 2011).

To date, the Global Runoff Data Base (GRDB) maintained by the Global Runoff Data Centre has been the primary dataset used in large-scale hydrological studies, with more than 9000 stations available to the research community (GRDC, 2015). The Global Runoff Data Centre (GRDC) database operates under the auspices of the UN – World Meteorological Organization (WMO), and its database is supported on a voluntary basis so that the number of data submissions depends on contributions by national authorities. However, although numerous countries have databases of acceptable quality, data supply remains resource intensive and the GRDB remains sparse in some regions. For example, the latest catalogue of the GRDB database (version 5 December 2017) shows that out of 7238 daily time series, there are only 637 stations over South America and only 642 stations over Asia. Moreover, many stations in regions such as Asia and Russia have not been updated for many years and are missing otherwise available data at the end of their records.

The Global Streamflow Indices and Metadata (GSIM) project has been initiated in order to address the demand for a global streamflow database (Bierkens, 2015; Fekete et al., 2015; Hannah et al., 2011; Kundzewicz et al., 2013; Merz et al., 2012; Milly et al., 2015). The approach of this project is not to collect high-quality data from referenced hydrological networks, which have been conducted in other studies (Addor et al., 2017; Burn et al., 2012; Hannaford and Marsh, 2006; Hodgkins et al., 2017; Whitfield et al., 2012) to support research that requires assumptions regarding the minimum impact of human interference on streamflow, such as the investigation of climate change implication for changes in extreme events. Instead, the activities of the GSIM project have been to collate publicly available data, apply basic consistency to the formatting, and establish a standardized set of metadata. In so doing, GSIM intends to promote more widespread use of streamflow data, facilitate improved research outcomes through increased spatial coverage and gauge density, and tackle ongoing challenges for the hydrological community, for example, addressing fundamental issues of data quality, identifying additional data sources, lobbying for continuity of data networks, and developing a method for improved governance and maintenance of streamflow data at the global scale.

To maximize the value of the streamflow dataset for a wide range of applications, the GSIM project also seeks to provide information on catchment characteristics upstream of the streamflow gauging station. This necessitates a consistent approach to delineating the upstream catchment boundary for every gauge station, and this is achieved using data from a global digital elevation model (DEM). This is because, with the exception of the GRDB databases, catchment boundaries representing the direct drainage area of stations were unavailable. Filling in this missing element of metadata is important to facilitate further analysis of the streamflow observations with respect to a wide and ever-increasing variety of spatial datasets. Although there have been previous efforts in producing catchment boundaries for a smaller number of stations (Addor et al., 2017; Arsenault et al., 2016; Lehner, 2012; Schaake et al., 2006), similar work at this magnitude has not been undertaken. This task is complicated by a lack of precision in the supplied geographic coordinates of a given site; for example, when a catchment boundary is extracted, the corresponding calculated area may not match the reported area of the catchment and a procedure for checking minor shifts in the coordinates is needed to improve identification of the likely catchment boundary. The quality of the delineated catchment boundary is also made available to GSIM users and should be considered prior to using this data product and any accompanying information.

The availability of catchment boundaries for each gauge enables the association of environmental variables with each gauge by extracting them from corresponding global-scale gridded products. As part of the GSIM project, a number of global data products are provided as an additional dataset so that a user can readily filter the GSIM dataset according to specific interests, for example, by climate type, soil type, land-use type, irrigation area, and population density. Other potential applications of this auxiliary information might include a comparison to a database of dams for identifying upstream impacts; to remotely sensed estimates of forest cover or urban extent for determining land-use change; to population demographics for improving estimates of flood exposure; and to hydrological model outputs for evaluating model performance.

Finally, to facilitate benefits of this project to the broader community, indices characterizing water-balance aspects, hydrological extremes, and features of the seasonal cycle have been derived from the GSIM time series and will be made publicly available. To ensure standardized quality for the derived indices, a quality control procedure coupling the information provided by data providers and a data-driven approach was also applied.

This is the first paper of a two-part series detailing the production of GSIM and corresponding data products. This paper outlines the provenance of daily streamflow time series (Sect. 2), procedures for reformatting and combining the time series (Sect. 3), the development of metadata associated with each gauge (Sect. 4), an overall summary of the GSIM time series and metadata (Sect. 5), and data availability (Sect. 6). As the time-series database cannot be made available online due to varieties of terms and conditions from data providers, the second paper in this series (Gudmundsson et al., 2018) is dedicated to the production of streamflow time-series indices, including (1) checks for data quality, (2) the production of streamflow time-series indices, and (3) homogeneity assessment of the derived indices.

2 Daily streamflow data and where to find them

GSIM is a compilation of 12 databases that have either open-access or restricted-access policies, and that collectively represent a total of 35 002 stations. The spatial distribution and the number of stations available in each database are illustrated in Fig. 1 (continental-scale figures are also provided as a Supplement). A summary of the data sources is also provided in Table 1 and detailed information on each database is elaborated upon in the following sections. The list of databases identified as part of GSIM is not exhaustive of all possible data sources, only of those that were known to the authors and readily accessible within the project time frame. Where additional data are available in a convenient format, it may be possible to further augment GSIM in the future.

The various data sources were classified as either a "research database" or a "national database". The reasons for this classification are further outlined in Sect. 3, but relate to issues when merging databases and removing duplicate gauges. The data sources include the following.

1. Research databases: databases with daily streamflow data that have been compiled on an ad hoc basis from a variety of original sources by research organizations. This category includes five different databases: the Global Runoff Data Base (GRDB); the European Water Archive (EWA); the China Hydrological Data Project (CHDP) data archive; the GEWEX Asian Monsoon Experiment – Tropics (GAME) data archive; and the Regional Hydrographic Data Network for the Arctic Region (ARCTICNET) data archive.

2. National databases: databases with daily streamflow data made publicly available by national water authorities as part of water-related regulations. This category includes seven databases: the USGS Water Data for USA database (USGS); Canada's National water data archive (HYDAT); Japan's Ministry of Land and Infrastructure database Water Information System (MLIT); Spain's digital hydrological yearbook database (Anuario de aforos digital 2010–2011, AFD); Australia's Bureau of Meteorological Water Data Online database (BOM); India's Water Resources Information System database (WRIS); and Brazil's National Water Agency database (ANA).

2.1 The Global Runoff Data Base (GRDB)

The daily streamflow dataset received from the GRDC (6313 stations with more than 10 years on record; see also Gudmundsson and Seneviratne, 2016) is referred to as the GRDB in this project. To date, the GRDB has been the largest and most extensively used dataset for streamflow analysis at regional and global scales. It was thus considered as the starting point and "base" for the GSIM project. Indeed, it was awareness of data not available from the GRDB that prompted the initial search for additional sources of data to complement the database.

The GRDC was initiated in 1988 by the WMO and is now maintained at the German Federal Institute of Hydrology in Koblenz. The GRDC provides free and unrestricted access to all hydrological data and products, although the data policy indicates that requests for data must reach the GRDC in written form to ensure data users do not redistribute the time series. More detail about the GRDC data policy, and the procedure for obtaining its time series, are outlined at http://www.bafg.de/GRDC/EN/01_GRDC/12_plcy/data_policy_node.html (last access: 23 June 2017).

2.2 The European Water Archive

The European Water Archive (referred to as the EWA in this paper) is one of the most comprehensive streamflow time-series archives in Europe, with more than 3000 river gauging stations distributed across 29 countries. This archive is also currently held by the GRDC and available under the GRDC data policy (http://www.bafg.de/GRDC/EN/04_spcldtbss/42_EWA/ewa_node.html, last access: 3 January 2018). The EWA stations used in this paper were selected using the same criteria as Gudmundsson and Seneviratne (2016), with a total of 3731 daily records.

2.3 The China Hydrology Data Project

The China Hydrology Data Project (CHDP) aims to digitize an arrangement of hydrological measurements taken at Chinese stations. These measurements (including daily discharge) were originally only available in book form (Henck et al., 2010). The original data were collected by the Chinese Hydrology Bureau and published in annual yearbooks. At the time GSIM began, discharge data were only available for the Yunnan-Tibet International Rivers, which corresponded to 163 stations until 1987. This project has been terminated since the 2000s and thus no further update is available. The data and metadata were obtained directly from the author of the project. Detailed information can be viewed at http://www.oberlin.edu/faculty/aschmidt/chdp/index.html (last access: 23 June 2017).

2.4 The GEWEX Asian Monsoon Experiment – Tropics project

The GEWEX Asian Monsoon Experiment – Tropics project (GAME) was initiated in 1996 to monitor several hydroclimatological variables over the humid temperate area in South-East Asia. As one of several important activities in this project, many hydrological observation datasets were collected, including streamflow data. Available streamflow data were provided by the Royal Irrigation Department of Thailand, and comprised 129 time series spanning a period from 1980 to 2000. Daily discharge data and associ-

Table 1. Basic information of daily streamflow databases included in the GSIM project.

Database (referred name)	Database category	Spatial coverage	Data access information
Global Runoff Data Base (GRDB)	Research database	Global	www.bafg.de/GRDC/ (last access: 23 June 2017) Archived database can be obtained via written request to the Global Runoff Data Centre. This database is updated when new data are submitted by national suppliers.
European Flow Regimes from International Experimental and Network Data (EWA)	Research database	European	http://ne-friend.bafg.de/servlet/is/7413/ (last access: 23 June 2017) Data can be obtained via written request to the Global Runoff Data Centre. This database has been frozen since October 2014 and is being integrated into the GRDB database.
A Regional, Electronic, Hydrographic Data Network for Russia (ARCTICNET)	Research database	Russia	http://www.russia-arcticnet.sr.unh.edu/ (last access: 23 June 2017) Archived and closed historic database. Part of this data archive has been included in the databases of the Global Runoff Data Centre and updated based on data deliveries.
China Hydrology Database Project (CHDP)	Research database	China	http://www.oberlin.edu/faculty/aschmidt (last access: 23 June 2017) Archived and closed historic database can be obtained via written request to the author of the database.
GEOSS ana MAHASRI Experiment in Tropics (GAME)	Research database	Thailand	http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/routine/rid-river/disc_d.html (last access: 23 June 2017) Archived and closed historic database
US National Water Information System (USGS)	National database	USA	http://waterdata.usgs.gov/nwis (last access: 23 June 2017) Individual time series can be downloaded from the data portal (updated regularly).
Canada National Water Data Archive (HYDAT)	National database	Canada	https://ec.gc.ca/rhc-wsc/ (last access: 23 June 2017) Archived database. The archive is updated quarterly by the data authority.
Brazil National Water Agency (ANA)	National database	Brazil	http://hidroweb.ana.gov.br/ (last access: 23 June 2017) Individual time series can be downloaded from the data portal (updated regularly).
Japan Water Information System (MLIT)	National database	Japan	http://www1.river.go.jp/ (last access: 23 June 2017) Individual time series can be downloaded from the data portal (updated regularly).
Anuario de aforos digital 2010–2011 (AFD)	National database	Spain	http://ceh-flumen64.cedex.es/anuarioaforos (last access: 23 June 2017) Archived database, DVD available from Spanish authorities (updated annually)
Australia Water Data Online (BOM)	National database	Australia	http://www.bom.gov.au/waterdata/ (last access: 23 June 2017) Individual time series can be downloaded from the data portal (updated regularly).
Water Resources Information System of India (I-WRIS)	National database	India	http://www.india-wris.nrsc.gov.in/wris.html (last access: 23 June 2017) Individual time series can be downloaded from the data portal (updated regularly).

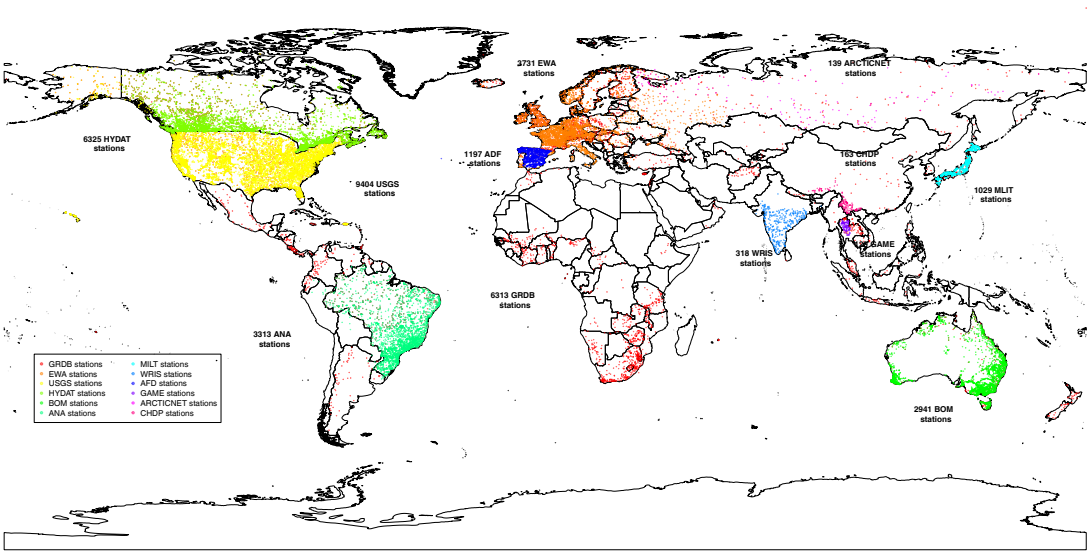


Figure 1. The distribution of stations from original data sources.

ated metadata were archived and can be accessed online at http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/routine/rid-river/index.html (last access: 23 June 2017).

2.5 The ARCTICNET project

A regional hydrometeorological data network for the pan-Arctic Region project is a regional database that can be accessed via the Internet and is referred to as ARCTICNET in this paper. The database is designed to support hydrological sciences and water resource assessments over this region with the goal of estimating the contemporary water and constituent balances for the pan-Arctic drainage system. ARCTICNET is a static dataset and some time series have been included in the databases of the GRDC and updated based on data deliveries. Although most data provided in the data portal are at monthly resolution, there are 139 high-quality daily streamflow time series across Russia that are also available which have not been fully integrated into GRDB. Although ARCTICNET's future status is likely to be a part of the GRDB, these stations have still been considered in GSIM production and are referred to as the ARCTICNET database in this paper. These time series, along with their metadata, were archived and can be downloaded at http://www.r-arcticnet.sr.unh.edu/v4.0/index.html (last access: 23 June 2017).

2.6 The USGS database

The USGS National Data Services for the US provide access to water resources data collected at approximately 1.5 million sites in all 50 states of the USA, also including the District of Columbia, Puerto Rico, the Virgin Islands, Guam, American Samoa, and the Commonwealth of the Northern Mariana Islands. All time series and associated metadata can be queried from the data portal http://waterdata.usgs.gov/nwis (last access: 23 June 2017). To ensure the queried data have sufficient geographic metadata (critical for the present project), the stations listed in the Geospatial Attributes of Gages for Evaluating Streamflow, version II (GAGES II) database were used (Falcone, 2011). The time series from 9404 stream gauges obtained from the USGS data portal are referred to as the USGS database in this paper.

2.7 The HYDAT database

Canada's National Water Data Archive (HYDAT) is a database containing daily observed hydrometric data from publicly funded gauges in Canada. Also available in the HYDAT database are metadata about the hydrometric stations, such as latitude and longitude, catchment area, record length, as well as information regarding flow conditions (current status, regulated or natural regime). The database is updated four times per year and currently contains data for 6325 streamflow stations across Canada. The raw data, as well as an extractor executable, are publicly available from Environment Canada's website at https://ec.gc.ca/rhc-wsc/default.asp?lang=En&n=9018B5EC-1 (last access: 23 June 2017).

2.8 The ANA database

The HIDROWEB data portal was organized by the Brazilian National Water Agency (ANA). It provides a database with all the information collected by Brazil's hydrometeorological network. Streamflow data and associated metadata were made publicly available by Brazil's national water regulations, and have been used extensively to monitor critical

events, such as floods and droughts. Individual time series and their associated metadata can be viewed or downloaded at http://hidroweb.ana.gov.br (last access: 23 June 2017). The 3313 stations downloaded from this website are referred to as the ANA in this paper.

2.9 The AFD database

Spanish streamflow data were retrieved from the digital hydrological yearbook (Anuario de aforos digital 2010–2011, AFD), which provides observations until 2013–2014 and is freely accessible online at http://ceh-flumen64.cedex.es/anuarioaforos/default.asp (last access: 23 June 2017). For the GSIM, we used the time series that was used to develop the E-RUN dataset (Gudmundsson and Seneviratne, 2016). The original DVD containing the full database was obtained directly from the Spanish authorities via a written form request. This collection contains streamflow data from 1197 gauging stations, and is referred to as ADF in this paper.

2.10 The MLIT database

In Japan, the Ministry of Land, Infrastructure, Transport and Tourism is responsible for organizing hydrological data. All records are disseminated at http://www1.river.go.jp/ (last access: 23 June 2017). As of 2010, the database kept records of all river stations (at both discharge and gauge level). The composition of the 15-digit station IDs is outlined in the file http://www1.river.go.jp/kitei_sosoku.pdf (PDF), and can be used to query and download time series, along with its metadata. As the whole database is recorded in Japanese, the translateR package (Lucas and Tingley, 2016) was used to translate the metadata into English. The time series downloaded from the Japanese water data portal (1029 stations in total) is referred to as MLIT in this paper.

2.11 The BOM database

As part of the water reform programme established in Australia, Water Data Online was created to provide free access to nationally consistent, current and historical water information. It can be accessed at http://www.bom.gov.au/waterdata (last access: 23 June 2017). Water Data Online also contains historical data from some stations that are no longer operational. Users can view or download individual streamflow time series from the data portal, along with standardized data and reports. The time series measured at 2941 stations obtained from Water Data Online is referred to as the BOM database in this project.

2.12 The WRIS database

The Generation of Database and Implementation of Web Enabled Water Resources Information System in the Country project (India-WRIS WebGIS) was initiated as a joint venture of the Indian Central Water Commission (CWC) and the Indian Space Research Organization (ISRO). Unclassified data can be accessed online and free of charge at http://www.india-wris.nrsc.gov.in/wris.html (last access: 23 June 2017), while the metadata are documented at http://www.cwc.nic.in/main/downloads/HydrologicalnetworkdetailsofCWC.pdf (last access: 23 June 2017). All 318 stations were downloaded from the website. They are referred to as the WRIS database in this paper.

The production of time series and metadata for GSIM comprises several stages due to the range of data formats and significant variation in the quality of metadata across data sources. To ensure GSIM is presented in a transparent manner, the following sections outline procedures that are used to collate the time series across (Sect. 3), and to produce the metadata (Sect. 4).

3 Procedure for combining databases

Several of the identified data sources share common spatial domains, where typically the research databases may contain a subset of gauges from the national databases. It is therefore important to correctly identify duplicate time series when merging the databases. To maximize the quality of combined time series and minimize the requirement to combine time series, this task is conducted following three sequential steps: Step 1 – pre-processing the data to a common structure; Step 2 – replacing all GRDB stations in countries that have a national database; and Step 3 – identifying remaining duplicates. From the 35 002 gauges, 3197 (2958 and 239 gauges from the GRDB and EWA databases, respectively) were replaced by national databases in Step 2, and 846 cases of "very likely identical" stations were identified and removed in Step 3, leaving 30 959 "duplication-free" time series available in the GSIM.

3.1 Pre-processing the time series into a singular data structure

One of the major challenges in producing consistent streamflow indices is that data from different sources have different structures and storage formats. For example, the MLIT database divides streamflow records at one location into separate text files, and each file contains streamflow measurements for 1 year. In comparison, the HYDAT archive includes streamflow measurements from all available stations in a single matrix.

To address the varying standards of data management, the first step in combining the databases was to reformat all the streamflow records to ensure that each time series is kept in a consistent format. Using the GRDB as a guide, it was decided to store all data for a given site in a single text file with three columns: (a) date of measurement, (b) value of measurement, and (c) original quality flags (if available), and with basic metadata (station name, ID, etc.) stored in the header of the

file. All additionally derived metadata (i.e. from global gridded products) are stored in the station catalogue. The streamflow measurements were also converted into consistent units (cubic metres per second).

Metadata that have special characters in foreign language sources were also pre-processed into the ASCII encoding system. For river names and station names that are recorded in Spanish (ADF) or Portuguese (ANA), the special characters were replaced by plain alphabetic characters using the core function iconv() of the R programming language. For river names and station names that are recorded in Japanese characters (MLIT), R package "translateR" (Lucas and Tingley, 2016) was used with the Google Translate API for this task. Although there are some limitations related to this toolset (e.g. some Japanese characters remaining untranslated and requiring manual translation; inconsistency in the translated results using the same original Japanese characters), this option was chosen to enable an automated and expedient translation. As a result, any text-related metadata associated with Japanese stations should be treated with care.

3.2 Replace the GRDB stations with national databases, if applicable

The streamflow records hosted by the GRDC (the GRDB and EWA databases) are themselves originally provided by national water agencies, and have been undergone quality control procedures by the GRDC. In cases that the supplied data contain errors, the GRDC informs data suppliers to improve the quality of their database. In term of data availability, time series downloaded directly from the national data portal usually represents the latest version of streamflow observation, and thus it seemed appropriate to replace stations hosted by the GRDC for countries where an equivalent national database was available. While this approach is efficient, there is a potential downside of removing GRDB stations that were not otherwise present in the national data depositories, perhaps due to differences in maintenance of the databases. Nonetheless, the number of stations available in the GRDB and EWA databases is much lower than that available in national databases for all countries (see Table 2). As a result of this step, 2958 stations located in seven countries (Australia, Brazil, Canada, India, Japan, Spain, and the United States) were removed from the GRDB collection. In addition, 239 stations located in Spain were removed from the EWA archive.

3.3 Identify and remove duplicates in research databases

The method of de-duplicating time series involves identification of duplicates where two data sources have overlapping coverage and potential merging of two records at a duplicated site to create a unified record. The de-duplication step was generally undertaken between the GRDB and a "paired" dataset (e.g. GRDB and GAME). The only exceptions to this step are for GRDB, EWA, and ARCTICNET, as these three datasets share Russia as a common spatial domain.

Table 2. Number of stations in countries where national databases are available.

Country	Database		
	EWA	GRDB	National
Australia	–	358	2941 (BOM)
Brazil	–	439	3313 (ANA)
Canada	–	1029	6325 (HYDAT)
India	–	0	318 (WRIS)
Japan	–	151	1029 (MLIT)
Spain	239	0	1197 (ADF)
United States	–	981	9404 (USGS)

The techniques adopted for combining research databases were based on the de-duplication procedures developed in Gudmundsson and Seneviratne (2016), which consists of three sequential steps.

1. Identification of "duplication candidates" using metadata similarity. This step aims to identify time series with a high level of similarity in metadata (either within one database or across different databases). We used three similarity metrics to identify potential time series: (1) Jaro–Winkler distances, a metric representing the alphanumeric similarity of strings (Christen, 2012), applied to river names of two records; (2) Jaro–Winkler distances between station names of two records; and (3) geographical proximity estimated from geographical coordinates between two records. These metrics were normalized to have the same range between 0 and 1, where a value of 0 indicates identical metadata (e.g. the same geographic coordinates). This similarity analysis was run for each pair in the pool of stations, and any pair with an average value below 0.25 was identified as candidate duplicate records.

2. Classifications of duplication candidates using time-series similarity. This step aims to decide whether a specific pair of duplication candidates is likely to be identical. The overlapping period and correlation coefficient were used as criteria for making a decision. Firstly, all duplication candidates that do not share any overlap in their period of record are kept in the final GSIM collection, as they can represent separate time series even if they measured discharge at the same geographical location (e.g. due to reconstruction of the gauging station). Secondly, any time series with a correlation coefficient (R^2) lower than 0.90 was automatically identified as "very likely different" (26 pairs), whereas $R^2 > 0.99$ indicates "very likely identical" time series (786 pairs). Finally, candidates with $0.90 \leq R^2 \leq 0.99$

Table 3. Basic metadata available from data sources.

Database	Station ID	Station name	River name	Geographical coordinates	Station elevation	Drainage area	Catchment boundary
GRDB	X	X	X	X	X	X	X
EWA	X	X	X	X	X	X	X
CHDP	X	X	X	X	–	X	–
GAME	X	X	X	X	X	X	–
ARCTICNET	X	X	X	X	X	X	–
USGS	X	X	–	X	X	X	–
HYDAT	X	X	–	X	–	X	–
ANA	X	E	E	X	X	X	–
ADF	X	E	E	X	X	X	–
MLIT	X	E	E	X	–	–	–
BOM	X	X	–	X	–	–	–
WRIS	X	X	X	X	X	X	–

X: metadata available; –: metadata are unavailable; E: metadata are not available in English.

(65 pairs) were visually inspected and manually classified as "very likely identical" (60 pairs) or "very likely different" (five pairs). All time series in the "very likely different" category were retained while stations of the "very likely identical" category were processed using the de-duplication procedure (see below).

3. De-duplication of identical time series: regardless of whether identical time series come from either the same database or from different databases, records with the greater number of data points in the streamflow time series were kept while the other(s) were discarded. Although this approach has the downside of truncating the length of useful records, the number of time series that could be influenced by this approach is relatively low (846 time series, corresponding to 2.8 % of the total number of available streamflow records).

A visual example of the de-duplication procedure is provided in Fig. 2. The left panel demonstrates a case of "very likely identical" stations, when station number 2964035 in the GRDB database was identified as an identical gauge to W.16 in the GAME archive, based on the similarities between the provided metadata and correlation coefficient. The time series representing station "GAME_W.16" was kept in the final collection, while time series "GRDB_2964035" was removed. The right panel in Fig. 2 demonstrates a case of a "duplication candidate" with correlation coefficient of 0.92 (time series "GRDB_6123645" and "EWA_9110028"). These time series were visually inspected, assigned a "very likely different" label, and both time series were kept in the final collection.

4 Production of the GSIM metadata

Providing a consistent set of metadata for each site has been a significant undertaking for GSIM. This section outlines three main stages to developing the GSIM metadata: (1) consolidating all available basic metadata; (2) consistently delineating catchment boundaries for each site; and (3) developing a supplementary set of catchment-scale metadata based on the delineated boundaries.

4.1 Consolidating basic metadata from available sources

Following the GRDB format, each time series was accompanied by basic metadata, including

1. station ID,
2. station name,
3. river name of gauging location,
4. geographical coordinates of station,
5. elevation of station,
6. drainage area, and
7. catchment boundary from original data sources.

These data are useful for filtering stations according to specific criteria and analysis objectives. Moreover, the availability of a catchment boundary for the gauge enables additional catchment-scale metadata to be derived as necessary. However, not all of these basic metadata were available for all data sources. For example, the catchment boundary was only available for parts of the GRDB and EWA stations, the drainage area was unavailable in the BOM and MLIT databases, and though several data sources included river names in station names (BOM, HYDAT, USGS), these metadata were unavailable in English for other sources (MLIT, ANA, ADF). Table 3 further outlines the availability of basic metadata for each source.

The method for consolidating basic metadata for each station follows three steps.

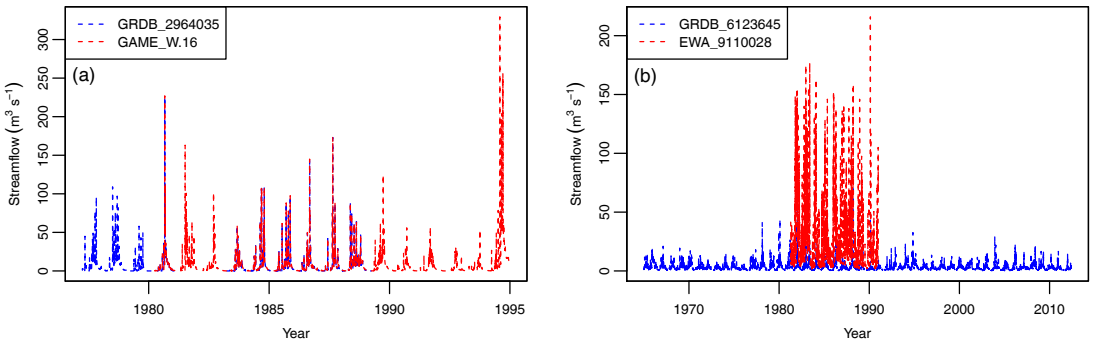


Figure 2. Examples of visually inspected duplication-candidate time series. **(a)** Two stations that were labelled "very likely identical" stations. **(b)** Two stations that were labelled "very likely different" stations.

Step 1. Transfer and review metadata available from original sources

The transfer of all existing metadata required a range of simple consistency checks and conforming rules, including the following.

1. Reviewing the geographical coordinates of all stations. Stations with unreasonable locations (e.g. located in the middle of the North Atlantic Ocean without any land mass, identified from Google Earth) were marked to be excluded from the subsequent delineation procedure (24 stations).

2. Separating the river name from the station name. Several sources use a consistent format for the station name consisting of two parts: the name of the station followed by the name of the water body. This pattern used a formula with "linking words" such as "at", "upstream" and "downstream". Taking station "BOM_406219" with original station name "Campaspe River at Lake Eppalock (Head Gauge)" as an example, the position of linking word "at" was identified and used to extract "river" metadata (Campaspe River) from the full station name.

3. Retaining the metadata of duplicated time series with the most data points in contrast to the other time series being removed. While this step may mistakenly remove some information, it is expedient and reflects the typical result of de-duplicated records that longer time series were kept while the shorter time series were removed.

Step 2. Generate "database-merging" information

This step documents a summary of efforts taken in creating a consistent set of GSIM metadata, and allows a user to check steps that were taken or to identify better procedures using alternative time series or metadata obtained from original sources. There are 12 fields documented for this purpose:

1. an indication of whether the time-series de-duplication procedure was used (one field),

2. which database and station were kept to construct the GSIM time series (two fields),

3. which station was removed and the corresponding database (three fields),

4. the value of metrics that represent similarities in the time-series metadata (five fields), and

5. the number of overlapping days, if applicable (one field).

Step 3. Generate information about data availability

The last step in compiling basic metadata for GSIM was to generate metrics that represent data availability for each GSIM time series, including the temporal coverage (i.e. the first and final years), the number of available daily observations, the number of missing data points, and the proportion of missing data points.

4.2 Catchment delineation procedure

With the ever-increasing availability of remote-sensing and modelled data products at global and continental scales, the provision of catchment boundaries is an important mechanism for extending the utility of GSIM. Although catchment boundaries can be generated easily using standard delineation algorithms in GIS packages, it requires a global coverage DEM dataset and reliable location to represent the outlet of each drainage area, which were unfortunately not readily available for GSIM project. This section describes the DEM products, and the algorithm to identify the "best outlet location" associated with each station that has been used in GSIM project.

The main DEM product used for GSIM was HydroSHEDS (http://hydrosheds.org, last access: 23 June 2017), which is

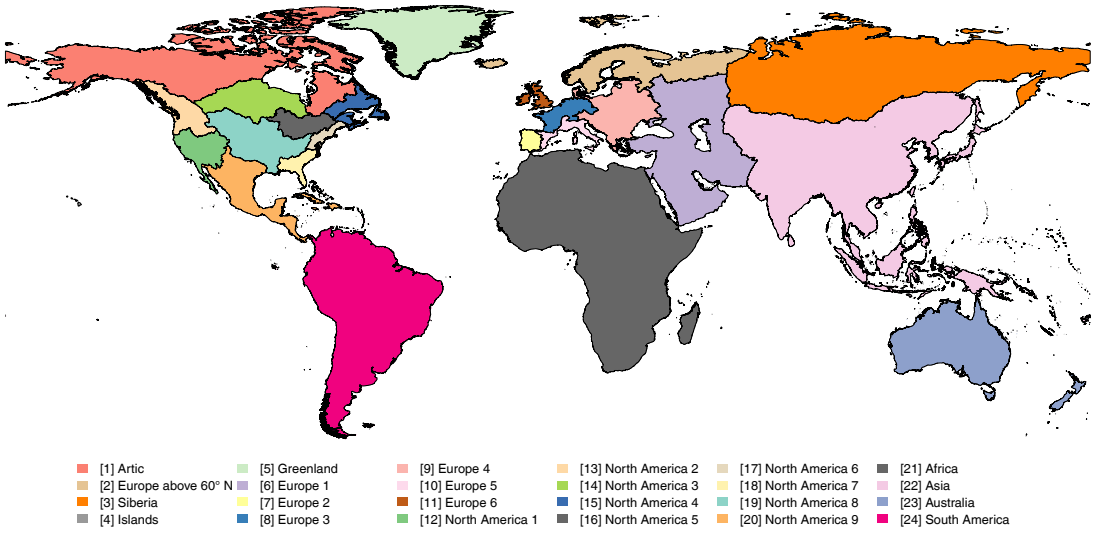


Figure 3. GSIM regions for catchment delineation and metadata extraction procedures.

available at 15 arcsec resolutions (Lehner et al., 2006), and has been used extensively in large-scale hydrological studies (Do et al., 2017; Lehner and Grill, 2013; Lehner et al., 2008; Wood et al., 2011). To address a limitation in the coverage of HydroSHEDS (no information in regions above 60° N, and some islands), the Viewfinder Panoramas elevation product at 15 arcsec resolutions was used (http://viewfinderpanoramas.org, last access: 25 June 2017) for those locations. This dataset has been used in several studies as an alternative DEM product to overcome similar data coverage issues (Barr and Clark, 2012; Fredin et al., 2012; Sil and Sitharam, 2016; Yamazaki et al., 2015). As there were more than 30 000 stations needing to be delineated, the HydroBASINS dataset was used, dividing the world into 24 regions, so that the task of delineation could be performed in parallel. The regions are shown in Fig. 3 and are generally independent in terms of drainage areas (Lehner and Grill, 2013). North America and Europe were specifically broken into more regions to address their relatively higher density of gauges. To maintain consistency when delineating boundaries, only one DEM product was used per GSIM region. As the quality of the Viewfinder Panoramas is not as clearly documented as for HydroSHEDS, its use was kept to a minimum. This resulted in five regions using Viewfinder DEM and 19 regions using HydroSHEDS (see Table 4).

Other challenges in the catchment delineation procedure are possible errors in the geographical coordinates representing the catchment outlet, such as typos in reported coordinates (e.g. 13.47° N instead of 14.47° N) or swapped order of the coordinate digits (e.g. 103.45° E instead of 103.54° E). These errors can lead to unreliable results of the delineation procedure, and so an algorithm to identify a location that represents catchment outlets well was also applied. This is described below.

Case 1. Reported station coordinates adopted as the outlet

If there was no information about a drainage area in the station metadata, the geographical coordinates of the station available from the data source were used as the outlet of the delineation process. There are automated techniques for repositioning outlets, such as choosing cells with the greatest flow accumulation within a search distance (Snap Pour Point ArcGIS tool), or finding the nearest cell possessing a flow-accumulation value above a specified threshold (Lindsay et al., 2008). Nonetheless, without information on the catchment area, it is impossible to assess the quality of the delineated catchment. Even if a repositioning technique were adopted, delineated catchment boundaries should be used with caution in this case, and therefore the original geographical coordinates was used to represent "best outlet location".

Case 2. Application of an automated repositioning algorithm

For stations with available information on catchment area, the automated repositioning procedure documented in GRDC report number 41 (Lehner, 2012) was used with some minor adjustments, and is summarized below.

1. The catchment area was estimated using the flow-accumulation dataset derived from the DEM products. This calculation was repeated for all pixels of the HydroSHEDS/Viewfinder gridded river network within a search radius of 5 km from the geographical coordinates of a specific station.

2. The estimated area values were compared with the reported area in the original metadata. All pixels were coded with the absolute value of their area differences

Table 4. DEM products used for each GSIM region.

Region	Description	DEM product
Arctic (region 1)	Represents the distant part of North America (including Alaska, most parts of Canada, and the eastern part of Autonomous Province, Russia)	Viewfinder DEM 15s
Europe above 60° N (region 2)	Represents countries located above 60° N (e.g. Sweden, Denmark, Norway, part of Germany, part of Russia)	Viewfinder DEM 15s
Siberia (region 3)	Represents areas above the 60° N part of Asia	Viewfinder DEM 15s
Islands (region 4)	Represents some islands across the Pacific Ocean (e.g. Honolulu, US) and Atlantic Ocean	Viewfinder DEM 15s
Greenland (region 5)	Represents land mass of Greenland	Viewfinder DEM 15s
Europe 1 to Europe 6 (six regions, from region 6 to region 11)	Represent most European countries (below 60° N)	HydroSHEDS DEM 15s
North America 1 to North America 9 (nine regions, from region 12 to region 20)	Represent US (except Alaska) and the southern part of Canada (below 60° N). It also includes central America for simplicity in processing catchment boundaries.	HydroSHEDS DEM 15s
Africa (region 21)	Represents Africa region	HydroSHEDS DEM 15s
Asia (region 22)	Represents Asia region (part of Kazakhstan, China, Mongolia, and Russia)	HydroSHEDS DEM 15s
Australia (region 23	Represents Australia, New Zealand, and some Pacific islands	HydroSHEDS DEM 15s
South America (region 24)	Represents South America	HydroSHEDS DEM 15s

(in %, with reported area in the metadata used as a reference). Pixels with area differences of more than 50 % were excluded. This procedure provided an area-based ranking scheme (RA) ranging from 0 to 50, where 0 indicates perfect agreement in catchment areas.

3. The distance to the original location of the station (geographical coordinates reported in the original metadata) was calculated for each pixel and normalized to reach 50 at the maximum distance of 5 km. This procedure provided a distance-based ranking scheme (RD) ranging from 0 to 50, where 0 indicates perfect agreement in station locations.

4. The final ranking scheme (R) was calculated as a combination of RA and RD, where distance rank was weighted twice as high ($R = \mathrm{RA} + 2\mathrm{RD}$) to penalize pixels that were further away from the original location.

5. The outlet was automatically relocated to the position of the pixel showing the lowest ranking value, and geographical coordinates of the pixel centroid were defined as the "best" outlet for this specific catchment.

6. In the original technical document (Lehner, 2012), a manual procedure was adopted for stations with differences in area above 50 (i.e. the search algorithm cannot find any pixel with an area difference less than 50 % within the 5 km search radius), or for stations that had no reported area in the data catalogue. This manual inspection process was infeasible given the scope of the GSIM project, having over 30 000 catchments being delineated and where river names were not available (or potentially inaccurately translated) for many stations.

A Python script was developed to automatically call the "best outlet location" algorithm and the catchment delineation toolset available in ArcGIS software (Jenson and Domingue, 1988) for each gauge using the chosen DEM data product. The delineated catchment boundary for each station was assigned a quality flag according to the discrepancy between reported drainage area and delineated catchment boundary area. There are four quality categories associated with the catchment boundary:

1. "High" quality: Area difference less than 5 %

2. "Medium" quality: Area difference from 5 % to less than 10 %

3. "Low" quality: Area difference from 10 % to less than 50 %

4. "Caution" quality: Area difference greater than or equal to 50 %, or the reported catchment area was not available in the GSIM catalogue.

Figure 4 demonstrates an example where the repositioning algorithm was used. Here the "best outlet location" was determined to be 4.8611 km away from the original location, which is defined by the reported geographical coordinates in the metadata (for station AR_0000007). The reported area in the metadata is 340 km^2, while the area of the delineated catchment boundary using the original coordinates was only 0.8 km^2, which is significantly lower than the correct number. On the other hand, the delineated catchment boundary using the "best outlet location" has an area of 363 km^2, indicating a better estimation of the upstream catchment boundary for this particular station.

4.3 Extraction of catchment-scale metadata

An important aspect of large-scale hydrology is the ability to exploit gridded datasets at the global scale (Bierkens, 2015; Bierkens et al., 2015; Gudmundsson and Seneviratne, 2015; Seneviratne et al., 2012; Ward et al., 2015). Having developed catchment boundaries for each GSIM station enabled a supplementary set of catchment-scale metadata to be derived with relative ease. A key feature is that the catchment boundaries and the subsequent metadata relates to the upstream contributing area that influences a gauge, rather than to the catchment (or arbitrarily defined sub-catchment) that contains the gauge and therefore includes a non-influencing downstream region.

In developing the catchment-scale metadata, a standard set of variables have been identified with a view to supporting a range of applications such as filtering stations according to characteristic features, performing analyses of streamflow according to explanatory features of a catchment, or classifying stations according to the (in)significance of human impact. As summarized in Table 5, a total of 12 global data products were used to derive 19 elements of catchment-scale metadata. These products were chosen to represent five main categories of catchment characteristics: (1) topography, (2) human impact, (3) climate type, (4) vegetation type, and (5) soil profile. Because the global data products have varying resolution and structure, the following method was used to derive the catchment-scale metadata.

1. Delineated catchment boundaries associated with each stream gauge were used to mask the subset of pixels from the resampled dataset.

2. If more than 30 % of the catchment area was not covered by a specific global data product, a "No data" code was given.

3. Metadata representing the characteristics of the upstream catchment for each streamflow gauge were calculated from the gridded data masked in step (1). There were three types of metrics calculated during this step.

   a. A single value. Used only for the elevation at the geographical coordinates of the gauge (i.e. the catchment outlet), number of large dams located within the catchment boundary, and total volume of corresponding reservoir.

   b. Average, min, max, and quartile values. Used for continuously varying data such as a slope or topography index. These metrics allow an idea of central tendency as well as spread of extracted data within each catchment boundary.

   c. Percentages of different classes of catchment characteristics. Used for categorical data. For example, there are 16 classes in the global lithology dataset, and the co-presence of more than one type of lithology occurs very often across all catchments. The percentages of each lithology class were therefore calculated and recorded for all available catchments. To make the results presentable in a final catchment-scale metadata matrix, an aggregated metric was calculated to indicate that there is a dominant class within the catchment boundary (i.e. more than 50 % of all available pixels). If there is no dominant class within the catchment boundary, a "No dominant class" string is provided.

5 Overview of the GSIM archive

This section summarizes the GSIM archive, including the availability of time series combined from 12 original data sources, the associated data products, and documentation outlining data quality (Sect. 5.1). The whole time-series database cannot be made available online due to data policies from a number of original data sources, some of which apply very strict terms and conditions regarding the redistribution of streamflow time series. To address this limitation and maintain the usefulness of GSIM to the research community, three metadata products have been developed and the availability of these data products is further discussed in Sect. 5.2.

5.1 Time-series availability

From the 35 002 time-series records obtained from 12 different sources, the final GSIM time-series archive holds a total of 30 959 unique stations, of which 30 935 stations have associated catchment shapefiles and catchment-scale metadata (24 stations were removed from this process due to suspect geographical locations). Most data sources are still active and being updated by the data authorities. GSIM, however, also included 425 "static" time series (from the ARCTICNET, GAME, and CHDP databases) that have been frozen since the early 2000s as these stations have improved the gauge

Table 5. Global data products used in GSIM and derived catchment-scale metadata.

Variables	Data sources	Spatial resolution	Reference period	Extracted metadata
Elevation	HydroSHEDS http://hydrosheds.org/ (last access: 23 June 2017) ViewFinder http://viewfinderpanoramas.org/ (last access: 23 June 2017)	15 arcsec × 15 arcsec	–	(1) Gauge elevation (2a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of catchment elevation
Slope	Derived from HydroSHEDS and ViewFinder DEM by authors	15 arcsec × 15 arcsec	–	(3a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of catchment slope
Topographic index	High-resolution global topographic index values (Marthews et al., 2015) https://catalogue.ceh.ac.uk/documents/ce391488-1b3c-4f82-9289-4beb8b8aa7da (last access: 23 June 2017)	15 arcsec × 15 arcsec	–	(4a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of catchment topographic index
Drainage density	GRIN – Global River Network (Schneider et al., 2017) https://www.metis.upmc.fr/fr/node/375 (last access: 23 June 2017)	7.5 arcmin × 7.5 arcmin	–	(5a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of catchment drainage density (km^{-1})
Dams	Global Reservoir and Dam (GRanD), version 1 (Lehner et al., 2011) http://sedac.ciesin.columbia.edu/data/set/grand-v1-dams-rev01 (last access: 23 June 2017)	6862 datapoints storage capacity of more than 0.1 km3	–	(6) Number of dams upstream (7) Total upstream storage volume
Population	Gridded Population of the World (GPW) version 4 (CIESIN, 2016) http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count (last access: 23 June 2017)	30 arcsec × 30 arcsec	2005–2014	(8a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of catchment population (2010) (9) 2010 Population count
Urbanization	Night Light Development Index (NLDI) dataset (Elvidge et al., 2012) http://www.soc-geogr.net/7/23/2012/sg-7-23-2012.html (last access: 23 June 2017)	0.25 arcdeg × 0.25 arcdeg	2006	(10a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of NLDI over catchment
Irrigation	Historical Irrigation Dataset (Siebert et al., 2015) https://mygeohub.org/publications/8/2 (last access: 23 June 2017)	5 arcmin × 5 arcmin	2005	(11a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of catchment Irrigated area (2005)
Climate type	World map of Koppen–Weiger climate classification system (Rubel and Kottek, 2010) http://koeppen-geiger.vu-wien.ac.at (last access: 23 June 2017)	5 arcmin × 5 arcmin	1951–2000	(12) Type of catchment climate (Koppen–Weiger) if one type present over more than 50 % catchment area, or "No dominant type"
Land cover	The Climate Change Initiative Land Cover (CCI-LC) dataset http://maps.elie.ucl.ac.be/CCI/viewer/download.php (last access: 23 June 2017)	7.5 arcsec × 7.5 arcsec	2015	(13) Type of catchment land cover (UN Land Cover Classification System) for 2015 if one type present over more than 50 % catchment area, or "No dominant type"
Lithological	The Global Lithological Map v1.0 (GLiM) dataset (Hartmann and Moosdorf, 2012) https://www.clisap.de/research/b:-climate-manifestations-and-impacts/crg-chemistry-of-natural-aqueous-solutions/global-lithological-map/ (last access: 23 June 2017)	0.5 arcdeg × 0.5 arcdeg	–	(14) Type of catchment lithology if one type present over more than 50 % catchment area or "No dominant type"
Soil profile	Soil grid 250 m (Hengl et al., 2017) https://soilgrids.org (last access: 23 June 2017)	7.5 arcsec × 7.5 arcsec	–	(15) Type of catchment soil class (World Reference Base) if one type present over more than 50 % catchment area or multiple types "No dominant type". (16a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of weight percentage of sand over the catchment (17a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of weight percentage of silt over the catchment (18a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of weight percentage of clay over the catchment (19a–f) Average, minimum, maximum, first quartile, second quartile, and third quartile values of bulk content of soil over the catchment ($kg\,m^{-3}$)

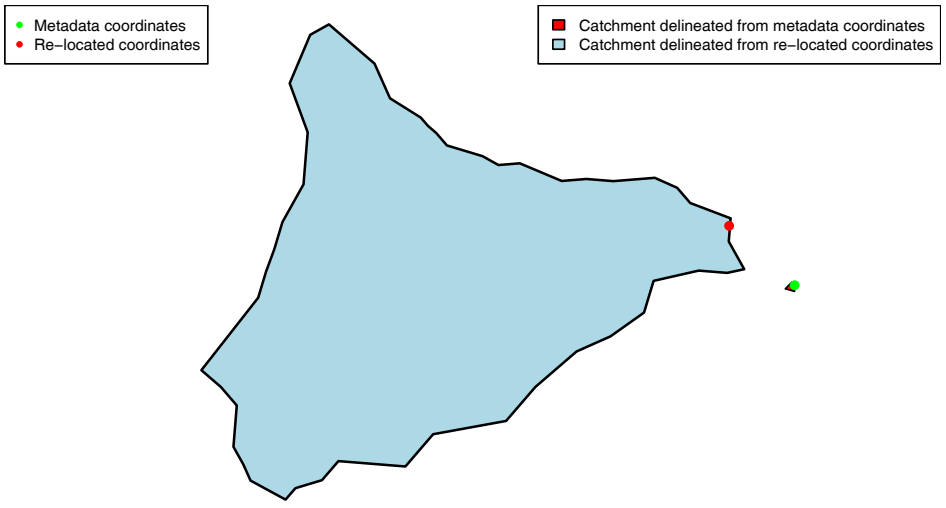


Figure 4. Example of improvement in quality of a catchment boundary using re-located geographical coordinates (for station AR_0000007).

density in regions with sparse streamflow observation systems (Russia, China, and Thailand, respectively). In addition, 2735 EWA stations (frozen since October 2014) were also included into GSIM as these time series have not been completely mirrored into GRDB database at the time GSIM was initiated. As these "static" time series have been frozen and no further update were provided, GSIM users are advised to use them with caution as the data may contain errors and/or have been replaced or updated.

As shown in Table 6, it is apparent that spatial coverage of the stations in the GSIM database varies significantly across continents, with North America and Europe having the greatest number of stations. Including the national databases such as MLIT (Japan), ANA (Brazil), BOM (Australia), and IWRIS (India) has significantly improved the observational network over the regions of Asia, South America, and Oceania (top panel of Fig. 5), some of which have recorded streamflow since the mid-20th century and were still operating at the time the GSIM database was initiated. This suggests that the national databases that are currently available should be given more attention in order to improve the quality and quantity of international archives.

Regarding temporal coverage, streamflow records across the globe are generally available for the second half of the 20th century (as shown in the bottom panel of Fig. 5). Regardless of missing data criteria, the number of available data gradually rises to its peak in the late 1970s to early 1980s, followed by a mild decrease in the late 1980s as also discussed by Hannah et al. (2011) and a secondary peak in the late 2000s. While the overall database has over 30 000 gauges, it is clear from Fig. 5 that from the 1960s onwards there are approximately from 10 000 to 15 000 gauges simultaneously active. This represents a significant increase in availability compared to the GRDB dataset, which had a total of approximately 9000 gauges and with a similar drop-off in available gauges depending on the filtering criteria applied.

5.2 Data products of GSIM

5.2.1 GSIM catalogue

The GSIM catalogue is designed for users to easily filter stations according to their purpose of application, and where necessary to transparently identify steps taken in the development of GSIM. The total number of 27 fields included in this document can be divided into three groups, namely the following.

1. Basic metadata. This group provides station identification, including a unique GSIM number, the name of the river, the name of the station, the elevation of the gauge, the provided geographical coordinates, and the catchment area.

2. Database merging metadata. This group of fields provides the identity of the numbers of original source(s), and if applicable the similarity metrics between duplicates.

3. Data availability metadata. This group of fields provides an overview of the data availability of each time series. These statistics were generated from the time-series data and can be used to filter station information, such as temporal coverage, data length, and the fraction of missing data.

As illustrated in Table 7, source datasets had significant gaps in the metadata, especially in cases of gauge elevation (not available in CHDP, GAME, HYDAT, BOM, and MLIT) and catchment area (not available in BOM and MLIT). In addition, the geographical coordinates of all stations were not

Table 6. Summary statistics of GSIM time series.

Continent	Number of stations	Average temporal coverage (years)	Shortest record (years)	Longest record (years)	Year of earliest entry	Year of latest entry
Africa	949	33.8	1	110	1903	2015
Europe	5778	40.3	1	208	1806	2016
Asia	1915	22.2	1	79	1921	2015
North America	15 884	42.9	1	156	1860	2016
South America	3449	29.3	1	116	1901	2016
Australia and Oceania	2984	31.4	1	131	1886	2016
Global	30 959	38.2	1	208	1806	2016

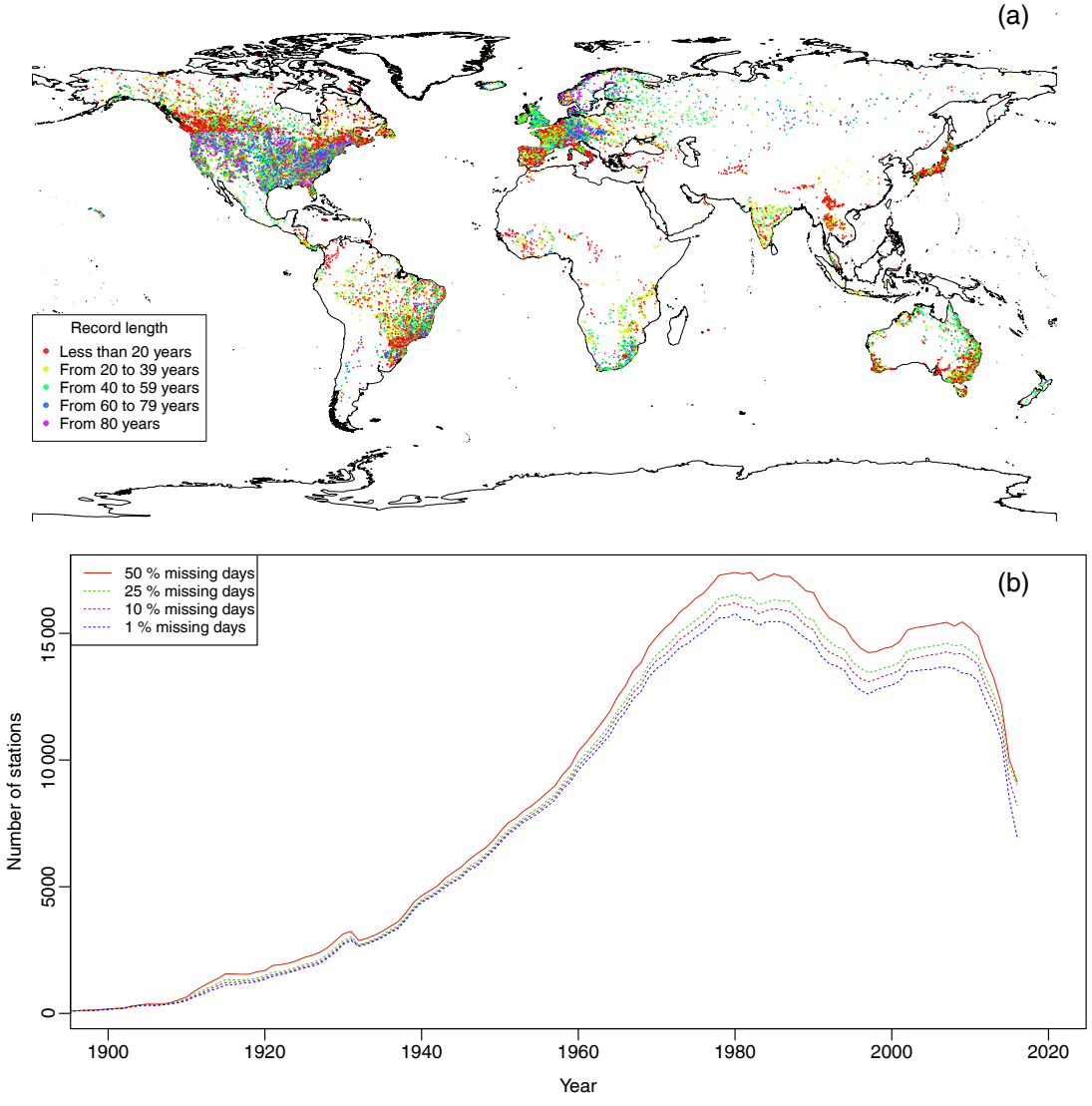


Figure 5. Availability of GSIM time series. **(a)** illustrates the length of record at each station, and **(b)** illustrates the number of available time series over time for four different missing data criteria.

Table 7. The percentage of stations accompanied by all basic metadata.

Dataset	Station ID	River name	Station name	Latitude	Longitude	Altitude	Catchment area
ADF	100	100	100	100	100	96.2	99.3
ANA	100	99.9	100	100	100	69	99
ARCTICNET	100	100	100	99.3	99.3	99.3	100
BOM	100	100	100	100	100	0	0
CHDP	100	99.4	100	100	100	0	84
EWA	100	100	100	100	100	98.5	94.5
GAME	100	100	100	100	100	0	100
GRDB	100	100	100	100	100	67	100
HYDAT	100	100	100	100	100	0	85.8
MLIT	100	100	100	100	100	0	0
USGS	100	100	100	100	100	93.7	25.5
WRIS	100	100	100	100	100	81.6	97.4
GSIM	100	99.9	100	99.9	99.9	50.4	73.8

Table 8. Percentages of available catchment-scale characteristics.

Catchment characteristics	Number of stations	Availability percentage
Climate classification	30 773	99.5
Drainage density	29 574	95.6
Elevation	30 932	99.9
Irrigation area	30 857	99.7
Land cover classification	30 888	99.8
Lithology type	30 154	97.5
Nightlight Development Index	23 096	74.7
Population count	30 894	99.9
Population density	30 800	99.6
Slope	30 862	99.8
Soil bulk density	30 812	99.6
Soil classification	30 764	99.4
Clay content	30 768	99.5
Clay content	30 695	99.2
Silt content	30 828	99.7
Topographic index	30 725	99.3

correctly recorded for all stations, with 24 removed as having suspect locations and 4871 shifted coordinates as part of the procedure for aligning catchment outlets with reported catchment areas.

5.2.2 Quality of catchment boundary

The catchment boundary is the second metadata product that is available through GSIM. Of all GSIM stations, 12 150 (39 %) were not associated with any information about drainage areas (including all MLIT and BOM stations); thus, a "Caution" flag is attached to upstream catchments of these stations. Another 24 stations with suspected geographical coordinates of stations were removed, and the final 18 785 stations were processed to identify the "best outlet" location to represent the outlet for delineating upstream catchments. The distribution and quality of the delineated catchments of these stations are provided in Fig. 6 (figures at continental scale are also provided as a Supplement).

As illustrated in the top panel, "Caution" catchments using "best" outlets (identified using the method outlined in Sect. 4.2) are generally located across all GSIM regions. However, the "Caution" flag appears more frequently over regions above 60° N. Further checks would be required to improve the association of catchment boundaries with stations. Unfortunately, the biggest caveat that applies to the GSIM database, as with any global database, is that the metadata were collated from a number of sources with varying standards of documentation and quality assurance and with limited capacity for additional checking other than automated procedures. Therefore, there is likely to be a non-trivial degree of error in the metadata for both geographical location and drainage area. Another issue that may lead to unreliable results of the delineation process is error in the DEM products. This potential error has been documented (Lehner, 2012; Lehner et al., 2006), and lower-quality DEM products generally exist for regions above 60° N due to the lower quality of the original elevation products used to derive the DEM datasets. Another note for the use of delineated catchments is that very small catchments (area less than 50 km^2) should be handled with care, as the "best" outlets could be located incorrectly while still delivering "acceptable" discrepancies as part of the automated procedure.

Nonetheless, the quality of delineated catchments is quite positive (as illustrated in the lower panels of Fig. 6). Of all 18 785 catchments that had reported drainage area in the GSIM catalogue, 68.25, 11.8, and 15.92 % of catchments have "High" quality (area discrepancy of less than 5 %), "Medium" quality (area discrepancy from 5 % to less than 10 %), and "Low" quality (area discrepancy from 10 to less than 50 %), respectively, while there are only 4.03 % catch-

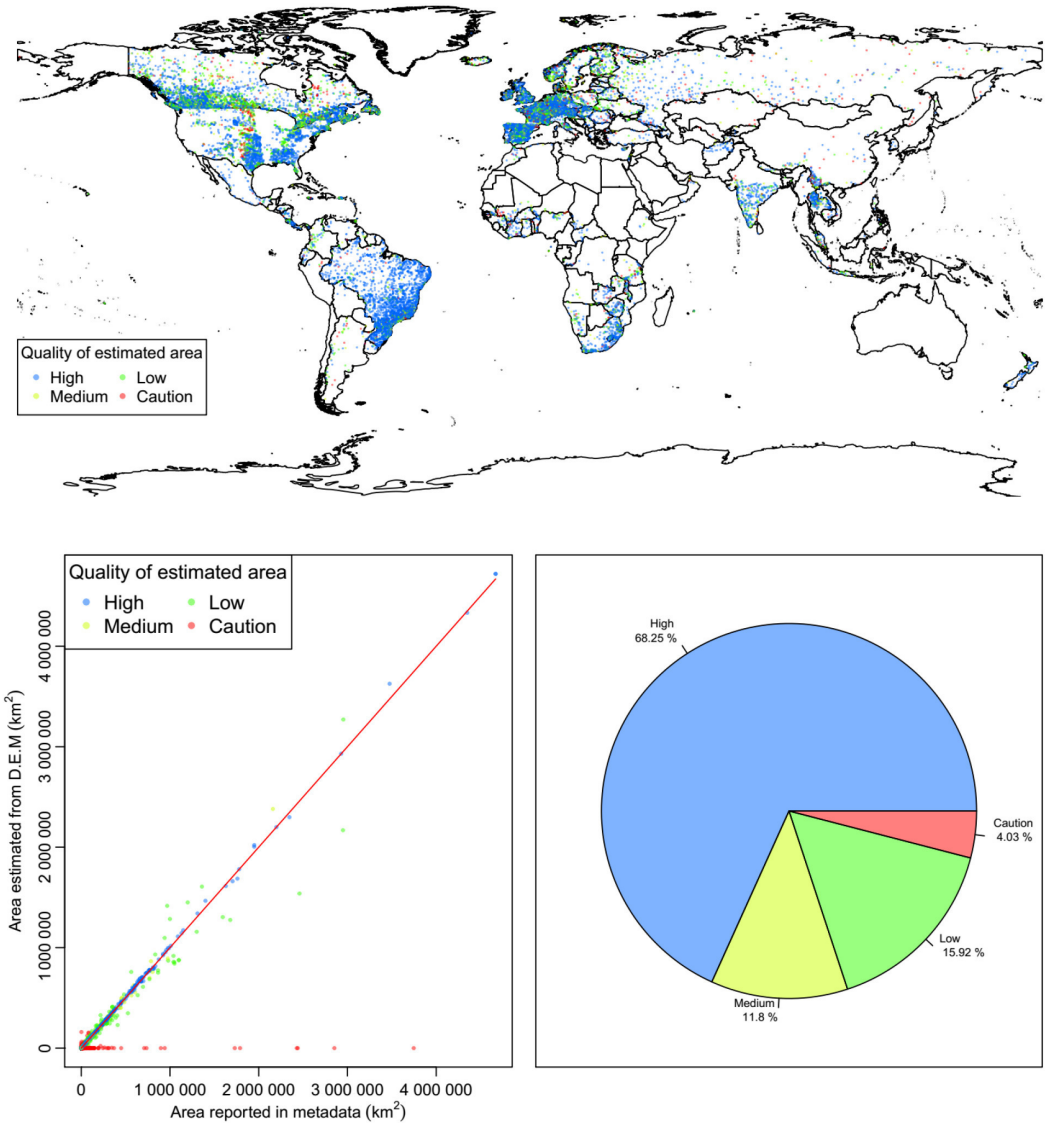


Figure 6. Quality of the delineated catchment boundary according to the categories of high, medium, low, and caution identified in Sect. 4.2 (for 18 785 stations that have reported drainage area and reasonable geographical coordinates).

ments with "Caution" quality (area discrepancy of more than or equal to 50 %).

5.2.3 Catchment-scale characteristics

The final data product that has been made available is the auxiliary information extracted from 12 global coverage datasets representing many characteristics associated with GSIM stations. Overall, the spatial coverage of original data products (mostly satellite-based is quite good (see Table 8), with just a small fraction of catchments (less than 10 %) that have more than 30 % of their areas not covered by these datasets. The exception is the Nightlight Development Index (NLDI – computed from the 2006 Nightlights dataset, Ziskin et al., 2010, and the 2006 Landscan gridded population, Bhaduri et al., 2002). This dataset does not have approximately 25.3 % of catchments covered, for more than 70 % of their areas.

It is important to note that while these catchment-scale characteristics are consistent products available for all stations, documentation for the original source data should be consulted during application to appreciate the limitations and appropriateness of each variable. For example, the GRanD database is not exhaustive of all dams worldwide and there can be ambiguities over the affiliated dates (e.g. whether they represent conception, construction, or commissioning). Furthermore, the extent of the overlapping period between temporal coverage of streamflow time series and remote sensing based datasets needs to be carefully assessed in cause–

effect studies. Similarly, it is likely that there will be updated or new data gridded datasets available over time so that applications should consider the appropriateness of the information used. The availability of metadata products emerging from the GSIM project demonstrates the possibility of using reported global data products to extract catchment-scale characteristics associated with each station with reasonable quality, enabling many potential applications from this rich information.

6 Data availability

The data described in this paper are available as a compressed zip archive containing (i) a readme file, (ii) metadata of all GSIM stations obtained from original data sources and time series, (iii) quality of catchment boundary and catchment characteristics extracted from 12 global data products, (iv) a list of stations with suspect geographical coordinates, and (v) catchment boundaries for 30 935 stations that have a reasonable geographical location.

The data can be freely downloaded at PANGEA data depository https://doi.pangaea.de/10.1594/PANGAEA.887477 (Do et al., 2018). The uploaded zip archive contains two directories and one README.txt file. The readme file provides a detailed description of the data. The "GSIM_catalogue" directory contains the metadata of all GSIM stations and a list of stations with suspect geographical coordinates. The "GSIM_catchments" directory contains shapefiles for 30 935 stations.

7 Conclusions

In situ observations of daily streamflow with global coverage are crucial to understanding large-scale freshwater resources that are fundamental for societal development. The GSIM archive, designed as an expansion of the GRDB database, has demonstrated the possibility of significantly improving the coverage and density of the global streamflow observational datasets using free-to-access databases. The development of the GSIM database was not possible without the tremendous investment in the production and ongoing maintenance of original data sources of GSIM. This fact emphasizes the key role of data authorities and international initiatives in enabling advances in large-scale hydrology by making data publicly available to the community.

While the activities of GSIM have been extensive in searching out and collating databases, they are by no means exhaustive (e.g. since submission we have been notified of additional potential candidates for inclusion such as the Mekong River Commission database, Chile national water database, and Argentina national water database). It is the authors' intention that this project will stimulate further efforts toward the development of coordinated and consistent representation of global streamflow observations. For this reason, the process of developing the archive was designed with automation in mind. With the exception of needing to visually inspect some cases of duplicated time series, the archive was automated using scripts in the R and Python programming languages.

Although the GSIM database was compiled from data sources that can be obtained free of charge via a data portal or by submitting written requests to data authorities, there are some strict conditions related to the redistribution of unprocessed data. Therefore, it is impossible to make the whole GSIM collection publicly available. In addition, with the main aim of harvesting as much data as possible, the GSIM database is not focused on collecting high-quality datasets such as referenced hydrological networks that are available in many countries (Whitfield et al., 2012), and thus the data quality may vary significantly across the available time series. To address these limitations and increase the usefulness of the GSIM database, we conducted a set of quality checking procedures for all GSIM time series. These quality-assured records were then used to produce a dedicated set of indices capturing important aspects of the daily dynamics from GSIM time series, and to explore potential applications of GSIM in large-scale hydrology. Detailed information about this work and associated distributed data is described in the second part of our series on GSIM (Gudmundsson et al., 2018a, b).

With the GSIM archive and production information made publicly available in a transparent manner, this project serves the broader hydrology community with improved coverage and quality of streamflow information. This project has yielded a significant increase in the availability of streamflow observations through the process of collating readily accessed online data, and with ongoing efforts there will be opportunities for further extension. Streamflow observations represent an underutilized resource, in part due to access limitations, but also due to challenges in accounting for human impacts in the observed record. These challenges notwithstanding, ongoing advances in global-scale hydrological models and ever-increasing access to remote-sensed products indicate that wider access to streamflow data has the potential to significantly enhance our knowledge of global water resources.

The Supplement related to this article is available online at https://doi.org/10.5194/essd-10-765-2018-supplement.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors would like to express their appreciation to all the national agencies and institutions that made the streamflow data available for this study. We would like to thank

Sonia I. Seneviratne for her discussions and support on the collation of the GSIM archive. Hong Xuan Do receives financial support from the Australia Award Scholarship (AAS). Seth Westra's time was supported by Australian Research Council Discovery project DP150100411. The authors also wish to thank two reviewers for their constructive comments and suggestions. The authors would like to express their sincere thanks to Danlu Guo for her support in collecting the MLIT database. This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

Edited by: David Carlson
Reviewed by: Wolfgang Grabs and one anonymous referee

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Arsenault, R., Bazile, R., Ouellet Dallaire, C., and Brissette, F.: CANOPEX: A Canadian hydrometeorological watershed database, Hydrol. Process., 30, 2734–2736, 2016.

Barr, I. D. and Clark, C. D.: An updated moraine map of Far NE Russia, J. Maps, 8, 431–436, 2012.

Bhaduri, B., Bright, E., Coleman, P., and Dobson, J.: LandScan, Geoinformatics, 5, 34–37, 2002.

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resour. Res., 51, 4923–4947, 2015.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrol. Process., 29, 310–320, 2015.

Burn, D. H., Hannaford, J., Hodgkins, G. A., Whitfield, P. H., Thorne, R., and Marsh, T.: Reference hydrologic networks II. Using reference hydrologic networks to assess climate-driven changes in streamflow, Hydrolog. Sci. J., 57, 1580–1593, 2012.

Christen, P.: Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection, Springer Science & Business Media, 2012.

CIESIN: Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals, NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, 2016.

Do, H. X., Westra, S., and Michael, L.: A global-scale investigation of trends in annual maximum streamflow, J. Hydrol., 552, 28–43, https://doi.org/10.1016/j.jhydrol.2017.06.015, 2017.

Do, H. X., Gudmundsson, L., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive – Part 1: Station catalog and Catchment boundary, PANGAEA, https://doi.org/10.1594/PANGAEA.887477, 2018.

Elvidge, C. D., Baugh, K. E., Anderson, S. J., Sutton, P. C., and Ghosh, T.: The Night Light Development Index (NLDI): a spatially explicit measure of human development from satellite data, Soc. Geogr., 7, 23–35, 2012.

Falcone, J. A.: GAGES-II: Geospatial attributes of gages for evaluating streamflow, US Geological Survey, 2011.

Fekete, B. M., Vörösmarty, C., and Grabs, W.: Global Composite Runoff Fields on Observed River Discharge and Simulated Water Balances/Water System Analysis Group. University of New Hampshire, and Global Runoff Data Centre, Koblenz, Federal Institute of Hydrology (BfG), Koblenz, Germany, Federal Institute of Hydrology (BfG), 2002a.

Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cy., 16, 15-11–15-10, 2002b.

Fekete, B. M., Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E., and Zhulidov, A. V.: Time for in situ renaissance, Science, 349, 685–686, 2015.

Fredin, O., Rubensdotter, L., van Welden, A., Larsen, E., and Lysså, A.: Distribution of ice marginal moraines in NW Russia, J. Maps, 8, 236–241, 2012.

GRDC: Report of the Twelfth Meeting of the GRDC Steering Committee, Koblenz, Germany, 18–19 June 2014, Global Runoff Data Centre (GRDC), Koblenz, Germany, 23 pp., 2015.

Gudmundsson, L. and Seneviratne, S. I.: Towards observation-based gridded runoff estimates for Europe, Hydrol. Earth Syst. Sci., 19, 2859–2879, https://doi.org/10.5194/hess-19-2859-2015, 2015.

Gudmundsson, L. and Seneviratne, S. I.: Observation-based gridded runoff estimates for Europe (E-RUN version 1.1), Earth Syst. Sci. Data, 8, 279–295, https://doi.org/10.5194/essd-8-279-2016, 2016.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe, J. Hydrometeorol., 13, 604–620, 2012.

Gudmundsson, L., Seneviratne, S. I., and Zhang, X.: Anthropogenic climate change detected in European renewable freshwater resources, Nat. Clim. Change, 7, 813–816, https://doi.org/10.1038/nclimate3416, 2017.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment, Earth Syst. Sci. Data, 10, 787–804, https://doi.org/10.5194/essd-10-787-2018, 2018a.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Time Series Indices and Homogeneity Assessment, PANGAEA, https://doi.org/10.1594/PANGAEA.887470, 2018b.

Hannaford, J. and Marsh, T.: An assessment of trends in UK runoff and low flows using a network of undisturbed catchments, Int. J. Climatol., 26, 1237–1253, 2006.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: importance, current status and future needs, Hydrol. Process., 25, 1191–1200, 2011.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochem. Geophy. Geosy., 13, Q12004, https://doi.org/10.1029/2012GC004370, 2012.

Henck, A. C., Montgomery, D. R., Huntington, K. W., and Liang, C.: Monsoon control of effective discharge, Yunnan and Tibet, Geology, 38, 975–978, 2010.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., and Bauer-Marschallinger, B.: SoilGrids250m: Global gridded soil information based on machine learning, PloS one, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., Fleig, A. K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C., and Wilson, D.: Climate-driven variability in the occurrence of major floods across North America and Europe, J. Hydrol., 552, 704–717, 2017.

Huang, S., Kumar, R., Flörke, M., Yang, T., Hundecha, Y., Kraft, P., Gao, C., Gelfan, A., Liersch, S., Lobanova, A., Strauch, M., van Ogtrop, F., Reinhardt, J., Haberlandt, U., and Krysanova, V.: Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide, Climatic Change, 141, 381–397, https://doi.org/10.1007/s10584-016-1841-8, 2016.

Jenson, S. K. and Domingue, J. O.: Extracting topographic structure from digital elevation data for geographic information-system analysis, Photogramm. Eng. Rem. S., 54, 1593–1600, 1988.

Kundzewicz, Z. W., Plate, E. J., Rodda, H. J., Rodda, J. C., Schellnhuber, H. J., and Strupczewski, W. G.: Changes in flood risk–setting the stage. In: Changes in flood risk in Europe, CRC Press, 2012.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G. R., Kron, W., Benito, G., Honda, Y., Takahashi, K., and Sherstyukov, B.: Flood risk and climate change: global and regional perspectives, Hydrolog. Sci. J., 59, 1–28, 2013.

Lehner, B.: Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network, Report 41 in the GRDC Report Series, available at: http://www.bafg.de/GRDC/EN/02_srvcs/24_rprtsrs/report_41.html?nn=201764 (last access: 23 June 2017), 2012.

Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrol. Process., 27, 2171–2186, 2013.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., and Magome, J.: Global reservoir and dam (grand) database, Technical Documentation, Version, 1, 2011.

Lehner, B., Verdin, K., and Jarvis, A.: HydroSHEDS technical documentation, version 1.0, World Wildlife Fund US, Washington, DC, 1–27, 2006.

Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, Eos, Transactions American Geophysical Union, 89, 93–94, 2008.

Lindsay, J. B., Rothwell, J. J., and Davies, H.: Mapping outlet points used for watershed delineation onto DEM-derived stream networks, Water Resour. Res., 44, W08442, https://doi.org/10.1029/2007WR006507, 2008.

Lucas, C. and Tingley, D.: translateR, available at: https://cran.r-project.org/web/packages/translateR/index.html (last access: 23 June 2017), 2016.

Marthews, T. R., Dadson, S. J., Lehner, B., Abele, S., and Gedney, N.: High-resolution global topographic index values for use in large-scale hydrological modelling, Hydrol. Earth Syst. Sci., 19, 91–104, https://doi.org/10.5194/hess-19-91-2015, 2015.

Merz, B., Kundzewicz, Z., Delgado, J., Hundecha, Y., and Kreibich, H.: Detection and attribution of changes in flood hazard and risk, Changes in flood risk in Europe, IAHS Special Publication, 10, 435–458, 2012.

Milly, P. C. D., Wetherald, R. T., Dunne, K., and Delworth, T. L.: Increasing risk of great floods in a changing climate, Nature, 415, 514–517, 2002.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., Stouffer, R. J., Dettinger, M. D., and Krysanova, V.: On Critiques of "Stationarity is Dead: Whither Water Management?", Water Resour. Res., 51, 7785–7789, 2015.

Nelson, B.: Data sharing: Empty archives, Nature, 461, 160–163, https://doi.org/10.1038/461160a, 2009.

Rubel, F. and Kottek, M.: Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification, Meteorol. Z., 19, 135–141, 2010.

Schaake, J., Cong, S., and Duan, Q.: The US MOPEX data set, IAHS publication, 307, 9–28, 2006.

Schneider, A., Jost, A., Coulon, C., Silvestre, M., Théry, S., and Ducharne, A.: Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density, Geophys. Res. Lett., 44, 2773–2781, 2017.

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., and Rahimi, M.: Changes in climate extremes and their impacts on the natural physical environment, Managing the risks of extreme events and disasters to advance climate change adaptation, 109–230, 2012.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., and Scanlon, B. R.: A global data set of the extent of irrigated land from 1900 to 2005, Hydrol. Earth Syst. Sci., 19, 1521–1545, https://doi.org/10.5194/hess-19-1521-2015, 2015.

Sil, A. and Sitharam, T.: Detection of Local Site Conditions in Tripura and Mizoram Using the Topographic Gradient Extracted from Remote Sensing Data and GIS Techniques, Nat. Hazards Rev., 18, 04016009, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000228, 2016.

Vörösmarty, C. J., Moore, B., Grace, A. L., Gildea, M. P., Melillo, J. M., Peterson, B. J., Rastetter, E. B., and Steudler, P. A.: Continental scale models of water balance and fluvial transport: an application to South America, Global Biogeochem. Cy., 3, 241–265, 1989.

Wanders, N. and Wada, Y.: Decadal predictability of river discharge with climate oscillations over the 20th and early 21st century, Geophys. Res. Lett., 42, 10689–610695, 2015.

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., van Beek, R., Bierkens, M. F., Ligtvoet, W., and Winsemius, H. C.: Assessing flood risk at the global scale: model setup, results, and sensitivity, Environ. Res. Lett., 8, 044019, https://doi.org/P10.1088/1748-9326/8/4/044019, 2013.

Ward, P. J., Eisner, S., Flörke, M., Dettinger, M. D., and Kummu, M.: Annual flood sensitivities to El Niño-Southern Oscilla-

tion at the global scale, Hydrol. Earth Syst. Sci., 18, 47–66, https://doi.org/10.5194/hess-18-47-2014, 2014.

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., Muis, S., de Perez, E. C., Rudari, R., Trigg, M. A., and Winsemius, H. C.: Usefulness and limitations of global flood risk models, Nat. Clim. Change, 5, 712–715, 2015.

Whitfield, P. H., Burn, D. H., Hannaford, J., Higgins, H., Hodgkins, G. A., Marsh, T., and Looser, U.: Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends, Hydrolog. Sci. J., 57, 1562–1579, 2012.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47, W05301, https://doi.org/10.1029/2010WR010090, 2011.

Yamazaki, D., Trigg, M. A., and Ikeshima, D.: Development of a global ∼ 90 m water body map using multi-temporal Landsat images, Remote Sens. Environ., 171, 337–351, 2015.

Ziskin, D., Baugh, K., Hsu, F.-C., and Elvidge, C. D.: Methods used for the 2006 radiance lights, Proceedings of the Asia-Pacific Advanced Network, 30, 131–142, 2010.

Appendix B - Paper 2 (constituting Chapter 3)

Do, H.X., Westra, S., Leonard, M. (2017) A global-scale investigation of trends in annual maximum streamflow, *Journal of Hydrology,* Volume 552, Pages 28-43, ISSN 0022-1694, https://doi.org/10.1016/j.jhydrol.2017.06.015.

Journal of Hydrology 552 (2017) 28–43

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

A global-scale investigation of trends in annual maximum streamflow

Hong X. Do *, Seth Westra, Michael Leonard

School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, South Australia 5005, Australia

ARTICLE INFO

Article history:
Received 13 September 2016
Received in revised form 2 June 2017
Accepted 10 June 2017
Available online 13 June 2017
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Saman Razavi, Associate Editor

Keywords:
Global flood hazard
Trend analysis
Climate change
Annual maximum streamflow

ABSTRACT

This study investigates the presence of trends in annual maximum daily streamflow data from the Global Runoff Data Centre database, which holds records of 9213 stations across the globe. The records were divided into three reference datasets representing different compromises between spatial coverage and minimum record length, followed by further filtering based on continent, Köppen-Weiger climate classification, presence of dams, forest cover changes and catchment size. Trends were evaluated using the Mann-Kendall nonparametric trend test at the 10% significance level, combined with a field significance test. The analysis found substantial differences between reference datasets in terms of the specific stations that exhibited significant increasing or decreasing trends, showing the need for careful construction of statistical methods. The results were more consistent at the continental scale, with decreasing trends for a large number of stations in western North America and the data-covered regions of Australia, and increasing trends in parts of Europe, eastern North America, parts of South America and southern Africa. Interestingly, neither the presence of dams nor changes in forest cover had a large effect on the trend results, but the catchment size was important, as catchments exhibiting increasing (decreasing) trends tended to be smaller (larger). Finally, there were more stations with significant decreasing trends than significant increasing trends across all the datasets analysed, indicating that limited evidence exists for the hypothesis that flood hazard is increasing when averaged across the data-covered regions of the globe.

1. Introduction

In recent decades, floods have caused nearly half of all weather-related disasters worldwide, and affected more than two billion people (CRED, 2015). The relative importance of flooding as a natural hazard has also increased over this period, whether measured in terms of economic losses (Kundzewicz et al., 2013), reinsurance losses (Mills, 2005) or the number of reported flood events (Munich Re, 2015; Swiss Re, 2015). Improved understanding of the causes of these changes is critical to manage and mitigate future impacts. However, the attribution of observed changes remains unclear, with possible causes including changes to the magnitude or frequency of high flow events (the flood hazard), the number of people or assets potentially affected by flooding (the flood exposure), the magnitude of impacts given a flood exposure (the flood vulnerability) (IPCC, 2012; Kron, 2005), or changes to reporting mechanisms and practices (Peduzzi et al., 2009).

One possible cause of the observed changes to flood impacts is the potential role of anthropogenic climate change. Recently, numerous studies have shown an intensification of extreme precipitation over data-covered land regions globally, and given that extreme precipitation is a leading cause of disastrous flooding (Guha-Sapir et al., 2015; Guha-Sapir et al., 2014), this intensification might cause an associated increase in flood hazard and thus flood impact. For example, Min et al. (2011) found that 65% of the data-covered areas of the globe exhibited increasing trends for annual maximum rainfall based on a gridded precipitation data product from 1951 to 1999. Using a point-based data record of 8326 high-quality land-based stations, Westra et al. (2013) detected increasing trends at nearly two-thirds of stations, and found that the median intensity of extreme precipitation increased in proportion to changes in global mean temperature at a rate of between 5.9% and 7.7% K^{-1}. Lehmann et al. (2015) also found large-scale increasing patterns in extreme precipitation, with 12% more record-breaking rainfall events over 1981–2010. This change was linked to an increasing trend in global temperatures over this time.

The observed intensification of extreme precipitation led the Intergovernmental Panel on Climate Change (IPCC) to find medium confidence for the conclusion that the increasing trends in extreme precipitation and associated discharge implies greater risk of

* Corresponding author.
E-mail address: hong.do@adelaide.edu.au (H.X. Do).

http://dx.doi.org/10.1016/j.jhydrol.2017.06.015
0022-1694/

flooding at the regional scale (IPCC, 2014). However, the implications of changes to extreme precipitation on discharge should be assessed with care due to the additional influence of a catchment's antecedent moisture content (i.e. the moisture stored in the catchment's soils, groundwater, lakes and reservoirs prior to the flood-producing rainfall event), which is affected by a catchment's long-term water balance (Johnson et al., 2016) rather than the intensity of individual heavy rainfall events. For example, only a third of discharge above the 99th percentile corresponded to precipitation above the 99th percentile (Ivancic and Shaw, 2015), indicating that the relationship between changes in extreme rainfall intensity and changes in flood hazard are complex and unlikely to be direct. This suggests that trends in extreme precipitation are not likely to be the only climatic factor influencing flood hazard, so that it is not possible to infer the direction and/or magnitude of change in flood hazard from information about changes in extreme precipitation alone.

An alternative approach to understanding changes in flood hazards at the global scale is through the use of large-scale hydrological models (Arnell and Gosling, 2014; Dankers et al., 2014). Studies that have taken this approach generally have found that more increases in flood hazard occurred in land grid cells than decreases, which is consistent with changes in climate variables generated from global climate models. However, high uncertainty remains a challenging issue in simulating streamflow at regional and local scales, as the direction of change is not always consistent across models for some river basins (Dankers et al., 2014). Although on-going efforts are being made to improve model resolution (Bierkens, 2015; Wood et al., 2011) and the representation of hydrologic processes (Clark et al., 2015), these models are still unable to represent many of the complex processes that are involved in the translation of rainfall to runoff at the global scale (Archfield et al., 2015; Sood and Smakhtin, 2015; Ward et al., 2015).

Rather than rely on indirectly inferring changes in flood hazard based on extreme precipitation data or using large-scale hydrological models to understand historical changes, several studies have directly analysed changes in streamflow data. However, challenges to this approach include data quality, availability, and the range of non-climatic factors that may also cause changes in flood hazard. In particular, compared to atmospheric variables such as temperature, pressure and rainfall, streamflow measurements are more susceptible to local (e.g. catchment-scale) anthropogenic influences. For example, land-use change, de-forestation, dams, reservoirs and other effects of urbanisation can all affect flood magnitude (Bradshaw et al., 2007; FitzHugh and Vogel, 2011), as can other hydraulic influences such as regulated water releases, changing channel capacity and/or implementation of flood prevention measures (Hewlett and Hibbert, 1967; Slater et al., 2015; Stover and Montgomery, 2001). There are also many practical challenges in characterising streamflow processes such as the difficulty of identifying subsurface contributions, inaccurate stage-discharge curves, tidal influences in estuarine catchments, and errors when measuring large flow events (Buschman et al., 2009; Di Baldassarre and Claps, 2011; Ghasemizade and Schirmer, 2013; Herschy, 1994). Finally, political sensitivity and costs associated with digitizing records or sharing data (Hannah et al., 2011; Nelson, 2009) can affect the availability of streamflow records, particularly at the continental or global scale.

Arguably as a result of the abovementioned challenges, studies that investigate changes in flood hazard directly based on observed flow data at the global scale are limited (Kundzewicz et al., 2004; Milly et al., 2002). In contrast to the extreme precipitation studies described earlier, these studies have relied on relatively small datasets (using 29 and 195 time series for the Kundzewicz et al. (2004) and Milly et al. (2002) studies, respectively). Furthermore, they did not lead to consistent conclusions regarding changes to flood hazard, with Kundzewicz et al. (2004) finding a similar number of stations exhibiting increasing and decreasing trends, and Milly et al. (2002) finding substantial increases in flood frequency for large basins.

In contrast to the limited number of global-scale studies, there have been numerous studies at regional scales, and some of these have been summarised in Table 1. However, it can be difficult to infer a picture at the global scale from these studies due to varying periods of data and strategies for selecting stations. Moreover, little effort has been made in relating any observed changes to natural or anthropogenic factors (Merz et al., 2012b) and there is an ongoing need for studies that use a consistent methodology for assessing streamflow across all regions to advance our knowledge of historical changes in floods (Kundzewicz et al., 2012; Merz et al., 2012a; Milly et al., 2015).

In order to better understand historical changes in flood hazard at the global scale, this paper therefore aims to improve the current understanding of changes to flood hazard by studying changes in the magnitude of annual maximum streamflow from a large global dataset of daily streamflow observations. The dataset is presented in Section 2, together with a number of catchment characteristics (climate region, catchment size, changes to forest cover and presence of large dams) that are used to screen the quality of records and investigate changes in streamflow with respect to potential influencing factors. The annual maximum daily streamflow and analysis methods are also described in Section 2. Trend results are presented in Section 3 with respect to the catchment and climatic characteristics of the streamflow location. Finally, implications for studies of streamflow observations and global-scale modelling efforts are discussed in Section 4.

2. Data and methodology

2.1. Overview of streamflow database

High-quality time series of streamflow are critical for change detection studies (Sheng and Linghui, 2012). However as indicated in the introduction, due to data limitations, most flood hazard change-detection studies are conducted at national and regional scales. These studies generally have relied on reference hydrometric network databases that have undergone extensive quality-checking procedures, and thus are less likely to be affected by low-quality time series (Whitfield et al., 2012). Furthermore, several of the datasets contain streamflow records only for catchments with minimum local anthropogenic influences (e.g. land use change, impoundments), so that any observed changes are more likely to be attributed by climatic causes (Burn et al., 2012). Examples of reference databases that have been widely used include the UK Benchmark Network (Hannaford and Marsh, 2006, 2008; Stahl et al., 2010); the Unites States Hydro-Climatic Data Network (Douglas et al., 2000; Lins and Slack, 1999; Mallakpour and Villarini, 2015; Schilling and Libra, 2003); the Canadian Reference Hydrometric Basin Network (Yue et al., 2003; Yue and Wang, 2002); and the Australia Hydrologic Reference Stations (Franks and Kuczera, 2002; Micevski et al., 2006; Turner et al., 2012).

In contrast to the above datasets, no reference database exists for global streamflow data, due to the diversity of instrumentation, collection and archiving methods used across different countries, and limitations in documentation (Hannah et al., 2011). The main dataset used for global-scale streamflow investigations is the Global Runoff Data Centre (GRDC) database, which was initiated in 1988 by the World Meteorological Organisation. The GRDC is maintained at the German Federal Institute of Hydrology in Koblenz (http://grdc.bafg.de), and holds records of 9213 stations

Table 1
Summary of observation-based studies on changes to flood frequency and/or magnitude.

Studies	Dataset	Scale of study	Main Findings
Milly et al. (2002)	29 daily time series spanning 1865–1999 from GRDC	Global	A substantial increase in the frequency of floods with discharges exceeding 100-year levels was identified from analysing annual maximum monthly-mean flows at 29 basins larger than 200,000 km^2
Kundzewicz et al. (2004)	195 daily time series spanning 1824–2002 from GRDC	Global	Annual maximum mean daily streamflows were analysed using the Mann-Kendall test at the 10% significance level. Only 27 (31) gauges had statistically significant increasing (decreasing) trends, and most (137) time series did not show any significant changes
Zhang et al. (2001)	249 Reference Hydrometric Basin Network (RHBN) records covering the 1947–1996 period	Canada	Annual maximum streamflows were analysed using the Mann-Kendall test at the 10% significance level, with significant decreasing trends in southern Canada and increasing trends in northern Canada
Cunderlik and Ouarda (2009)	169 stations in the RHBN with common period from 1974–2003	Canada	Seasonal maximum streamflows over the snowmelt period were analysed using the Mann-Kendall test at 10% significance level. Almost 20% of all stations showed significant trends in the magnitude of snowmelt floods. A notable finding is most of the detected significant trends were decreasing
Burn and Whitfield (2016)	280 stations obtained from Environment Canada Data Explorer database spanning the 1961–2010 period	Canada	Changes in the magnitude and timing of flood events exceeding the 25-, 50- and 100-year return periods were examined. The study found a generally decreasing flood magnitudes in nival catchments, and increasing flood magnitudes in pluvial catchments
Lins and Slack (1999)	395 time series covering the 1944–1993 period from U.S. Hydro-Climatic Data Network (HCDN)	U.S.	Annual maximum streamflows were analysed using the Mann-Kendall test at the 5% significance level. The study found a mix of increasing and decreasing trends in annual maximum streamflows across the eastern half of the United States, while trends in the western U.S. record mostly decreased
Douglas et al. (2000)	1474 time series spanning over 1874–1988 from HCDN	U.S.	Annual maximum daily streamflows were analysed using the Mann-Kendall test at the 5% significance level. Across U.S., no evidence of trends in flood flows was found
Mallakpour and Villarini (2015)	774 U.S. Geological Survey (USGS) stream gauges covering the common 1962–2011 period	Central U.S.	This study focused on assessing changes of flood magnitude and frequency in the Central United States using the Mann-Kendall test at the 5% significant level. The findings showed that flood frequency has increased while there was limited evidence of a decrease in flood magnitude in this region
Archfield et al. (2016)	345 USGS time series spanning the 1940–2013 period	U.S.	Four dimensions of floods (frequency, magnitude, duration and volume) were evaluated across U.S. Although detected trends were more than what would be expected by chance, this study could not identify any clear spatial pattern of changes in floods
Slater and Villarini (2016)	Daily gauge height data over 2042 catchments collected from USGS database	U.S.	The peak over threshold approach was used to identify the number of days exceeding the threshold of gauge height data. Trends were detected using a Poisson regression model at the 5% significance level. This study identified that there were increases in flood risk around the upper Midwest/Great Lakes region and decreases on the Gulf Coastal Plain, the southeastern United States, and California
Stahl et al. (2011)	441 time series spanning over 1962–2004 obtained from European Water Archive (EWA)	Europe	This study calculated the slope of Kendall-Theil robust line for 7-day annual maximum streamflows records and compared it with model simulation from global hydrology model. An increasing pattern in the north/west and decreasing pattern in the south/east of the European continent was found
Morán-Tejeda et al. (2012)	57 stations spanning the 1961–2005 period obtained from Duero water management agency	Spain	The 90th and 99th percentile of daily streamflow over a year was used to represent flood data. This study showed a general trend of decreasing frequency and magnitude of high flows throughout most of the basin
Bard et al. (2012)	342 AdaptAlp stations (obtained from seven data authorities) with at least 40 years of data records over the common period 1961–2005	Alps (Europe)	Snowmelt indices representing snowmelt streamflow intensity and seasonality were statistically investigated at the 10% significance level. This study showed an increasing trend in spring floods associated with snowmelt
Hannaford et al. (2013)	132 time series covering 1932–2004 period from EWA	Europe	Trends in 7-day annual maximum streamflow were indicated by the Mann-Kendall Z statistic. The results demonstrated that trends in flood magnitude are highly influenced by interdecadal variability.
Ishak et al. (2013)	330 stations spanning over 1955–2004 obtained from the Bureau of Meteorology	Australia	Annual maximum daily streamflows were analysed using the Mann-Kendall test at the 10% significance level. The assessment of trends indicated that the south-east and south-west regions of Australia have experienced a significant downward trend in the annual maximum streamflow over reference periods
Delgado et al. (2010)	4 stations spanning1913–2000 the Southern Institute of Water Resources Research – Vietnam	South-east Asia	Annual maximum daily streamflow were analysed using linear regression, the Mann-Kendall test at 10% significance level and a generalised extreme value model. The study found that the probability of average flood events has decreased during recent decades while extreme floods are likely to increase
Nka et al. (2015)	11 time series covering 1950–2010 period	Africa	Annual maximum daily streamflow were analysed using the Mann-Kendall test at 10% significance level. This study indicated that a mixture of both increasing and decreasing trends were found across West Africa

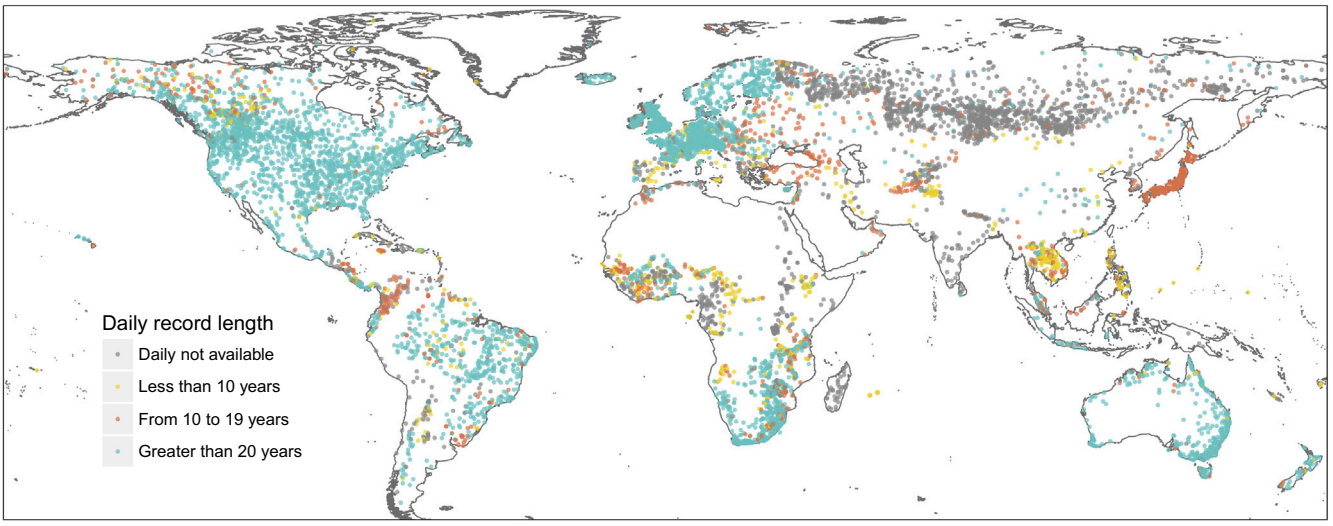


Fig. 1. Global coverage and record length of GRDC streamflow stations.

across the globe (spatial coverage shown in Fig. 1), with an average time series length of 42 years per station (GRDC, 2015). However, interpretation of any trends from the GRDC data requires caution since the catchments may or may not be affected by anthropogenic activities, and quality control processes are likely to vary depending on the country of origin.

2.2. Reference periods used for analysis

To facilitate the detection of change in flood magnitude, the minimum data record length should be longer than decadal periodicity of hydrological cycles (Sheng and Linghui, 2012). Of the 9213 records in the GRDC, the time series were initially filtered using the following procedure:

- For each station, 'flood' flows were obtained as the maximum daily value for each calendar year. In this step, the percentage of missing data was also estimated, and the annual maximum streamflow for that year was classified as 'missing' if more than 10% of days within that year were not available.
- Stations with less than 30 years of available annual maximum data (not necessarily continuous) were then removed (comprising 5620 stations in total). In addition, stations with instances of two or more identical annual maximum streamflow values were also identified and manually checked through visual inspection, leading to the removal of a further 35 stations due to suspect quality. These gauges were mostly located in Africa and South America.

The remaining 3558 stations represent the most comprehensive observation-based record of annual maximum streamflow at the global scale currently available. However, the number of annual maxima for these stations varies significantly through time (as shown in Fig. 2), reaching its peak in the late 1970s to 1990s followed by a decrease in the 2000s owing largely to data collation rather than discontinuation of gauges.

Given the changes in data availability with time, three separate reference periods were identified for this study:

- Dataset A1 (1907 stations) comprises stations with at least 38 years annual maximum streamflow records over the 1966–2005 period (average record length of 39.7 years). This strict admission criterion (no more than 5% missing data) helps ensure the consistency in identifying changes in flood hazard over the reference period (while not compromising spatial coverage) with a good balance between the length of data series and the number of stations.
- Dataset A2 (3478 stations) comprises stations with at least 30 years annual maximum streamflow over the 1955–2014 period (average record length of 47.6 years). This more generous admission criteria (up to 50% missing data is allowed) leads to a larger set of stations and a better spatial coverage. However, this dataset is also accompanied by a significant reduction in terms of data continuity.
- Dataset A3 (721 stations) comprises stations with at least 80 years annual maximum streamflow over the 1900–2014 period (average record length of 93.0 years). This long period together with fairly strict admission criteria (less than 30% missing data allowed) constrains the available stations to a smaller geographic area (largely North America and Europe) but with a better temporal coverage.

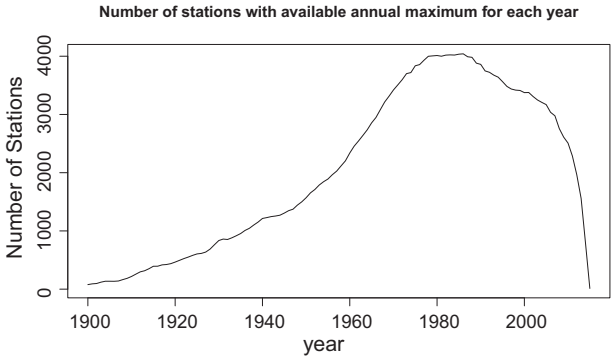


Fig. 2. The total number of stations with at least 30 years of available data that have suitable annual maximum streamflow records in any given year.

Of these three periods, dataset A1 represents the best compromise among data length, completeness and availability of stations. Therefore, this dataset is used as the primary dataset for subsequent analyses in this paper, with datasets A2 and A3 analysed for comparative purposes in Section 3.1.

2.3. Filtering streamflow data based on catchment characteristics

One challenge in detecting trends in annual maximum streamflow is that streamflow measurements are sensitive to local

anthropogenic influences (e.g. human regulation over the upstream catchment area). To reduce the influence of human-induced impact when evaluating trends, many studies have used catchment area as a selection criterion. Specifically, small catchments with areas less than 1000 km^2 have been preferred (Hannaford et al., 2013; Ishak et al., 2013; Stahl et al., 2010) as these catchments are considered less likely to have been extensively modified.

In this study, an alternative approach was taken whereby the upstream catchment area of each station was delineated and used to derive metadata such as the existence of large dams, forest-cover changes and climatological classification, which can be used to provide insight into the likelihood of anthropogenic changes in each catchment. To this end, catchment boundaries of the contributing upstream region for each gauge were obtained from the GRDC. A number of the GRDC streamflow records did not have catchments delineated, and for these catchments the HydroSHEDS flow direction data was used for the delineation (http://hydrosheds.cr.usgs.gov), which is available at 15 arc-seconds resolution (Lehner et al., 2006). The Viewfinder Panoramas elevation product (http://viewfinderpanoramas.org/) was also used as an alternative to address lack of coverage in the polar region and on some islands in the HydroSHEDS dataset. Having delineated each catchment, it is possible to associate catchment features with each streamflow gauge (see following sections for details).

2.3.1. The Global Reservoir and Dams (GRanD) database

The Global Reservoir and Dams (GRanD) database (Lehner et al., 2011) was used with delineated catchment boundaries to determine whether an upstream storage was present. The GRanD database contains storages with a capacity of at least 0.1 km^2, representing 6862 records of reservoirs and their associated dams. Although this is a valuable dataset for investigating the influence of dams on flood hazard, it is cautioned that the data is obtained from various research groups on a voluntary basis, and thus it is unlikely to provide a comprehensive description of all dams globally.

The catchment boundary positions of all 1907 stations in dataset A1 were compared with the dam coordinates to detect the presence of dam(s) and classify stations into two sub-datasets:

- Dataset B1 (1143 stations): Stations with no dams within the upstream boundary. This dataset is referred to as the "no dams" dataset.
- Dataset B2 (764 stations): Stations with at least one dam within the upstream boundary. This dataset is referred to as the "dams" dataset, which might be affected by dam construction or operation, leading to misinterpretation of detected changes.

2.3.2. Köppen-Weiger climate classification

World maps of the Köppen-Weiger climate classification (Rubel and Kottek, 2010) were used to identify the climate classification for each catchment. This dataset was developed using a gridded monthly temperature product provided by the Climatic Research Unit (CRU) of the University of East Anglia (Mitchell and Jones, 2005) and a gridded precipitation product (GPCC's full data reanalysis version 4) obtained from the Global Precipitation Climatology Centre (GPCC) for 1901–2007 (Fuchs et al., 2007). These observation-based datasets cover the global land areas excluding Greenland and Antarctica and have undergone a series of quality control steps to avoid inhomogeneities in the gridded data. The world map of Köppen-Weiger climate classification comprises 31 climate classes, which were adjusted into nine sub-climatic groupings and used in this study. The spatial distribution of the nine sub-climatic classifications is illustrated in Fig. 3.

The nine sub-climatic groupings were initially developed from the five major climatic groups. These major groups comprise equatorial climates (A), arid climates (B), warm temperate climates (C), snow climates (D) and polar climates (E). Across these five major groups, temperate climates (C) and snow climates (D) were regions with the highest density of available runoff stations, so that each of these classifications were further subdivided into three climate groups to create the nine-group classification system. The main features of the classified climates are provided in Table 2.

2.3.3. The Global Forest Change dataset 2000–2012

The Global Forest Change dataset 2000–2012 version 1.0 (http://earthenginepartners.appspot.com) was used to derive the proportions of change in forest area, in which "forest" is defined for locations with trees at least 5 m tall (Hansen et al., 2013). This

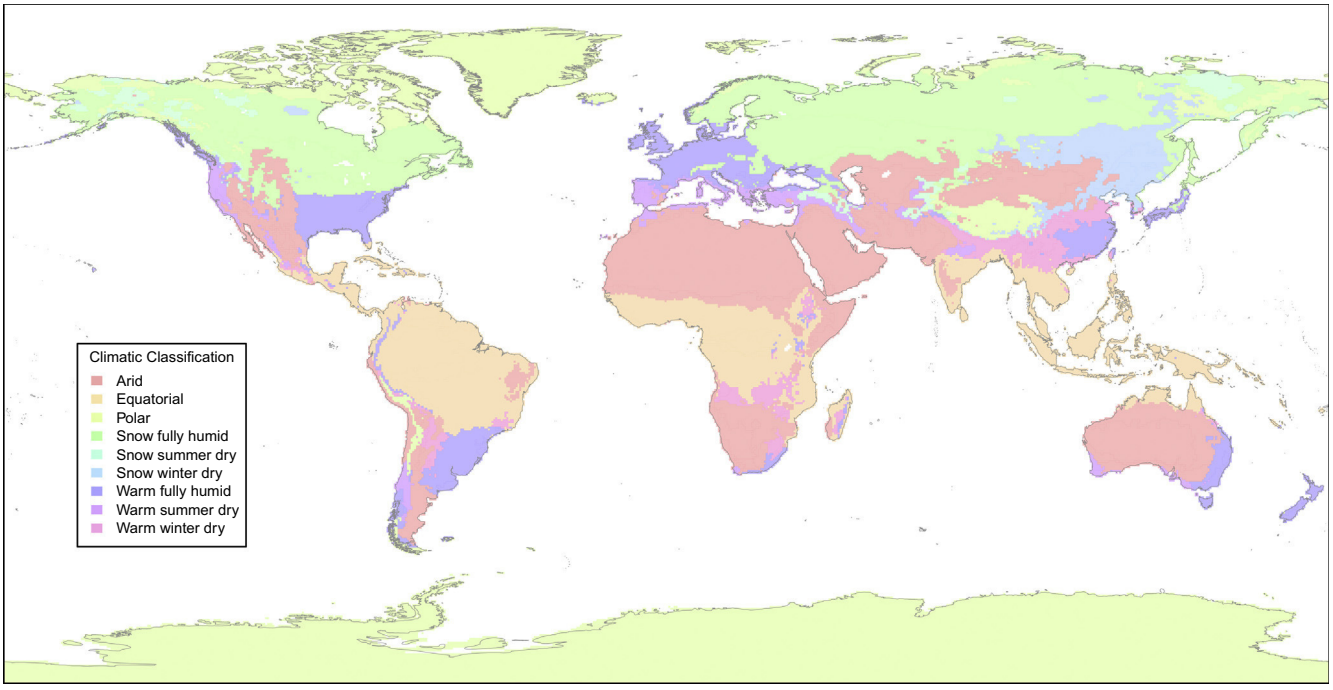


Fig. 3. Spatial distribution of nine groupings of Köppen climate types used in this study based on Rubel and Kottek (2010).

Table 2
Main features of nine climatic groups, based on the Köppen-Weiger climate classification system.

Climate groups	Description
Equatorial climates	$T_{min} \geq +18$ °C
Arid climates	$P_{ann} < 10\ P_{th}$
Warm temperate climates (C): Lowest monthly mean temperature between -3 °C and 18 °C	
Warm summer dry (Cs)	$P_{smin} < P_{wmin}$, $P_{wmax} > 3\ P_{smin}$ and $P_{smin} < 40$ mm
Warm winter dry (Cw)	$P_{wmin} < P_{smin}$ and $P_{smax} > 10\ P_{wmin}$
Warm fully humid (Cf)	Neither Cs nor Cw
Snow climates (D): Lowest monthly mean temperature ≤ -3 °C	
Snow summer dry (Ds)	$P_{smin} < P_{wmin}$, $P_{wmax} > 3\ P_{smin}$ and $P_{smin} < 40$ mm
Snow winter dry (Dw)	$P_{wmin} < P_{smin}$ and $P_{smax} > 10\ P_{wmin}$
Snow fully humid (Df)	Neither Ds nor Dw
Polar climates	$T_{max} \leq 10$ °C

T_{min} (°C): Lowest monthly mean air temperature.
T_{max} (°C): Highest monthly mean temperature.
P_{ann} (mm/year): Accumulated annual precipitation.
P_{wmin} (mm/month): Lowest monthly precipitation for the winter half-year.
P_{smin} (mm/month): Lowest monthly precipitation for the summer half-year.
P_{wmax} (mm/month): Highest monthly precipitation for the winter half-year.
P_{smax} (mm/month): Highest monthly precipitation for the summer half-year.
P_{th} (mm): Dryness threshold.
where: $P_{th} = \begin{cases} 2 * T_{ann} & \text{if at least 2/3 of annual precipitation occurs in winter,} \\ 2 * T_{ann} + 28 & \text{if at least 2/3 of the annual precipitation occurs in summer,} \\ 2 * T_{ann} + 14 & \text{otherwise.} \end{cases}$

satellite-derived product was obtained from Landsat 7 ETM+ images based on the Normalized Difference Vegetation Index (NDVI), and is available at one arc-second resolution. The Global Forest Change dataset comprises several products, and the following two layers were used in this study:

- Global forest cover loss 2000–2012 (loss): Forest loss during the period 2000–2012 was identified when a disturbance (or a change from a forest to a non-forest state) was detected. Each cell in this dataset was assigned either a value of 1 (loss) or 0 (no loss).
- Global forest cover gain 2000–2012 (gain): Forest gain during the period 2000–2012 was identified when the inverse of loss (or a change from non-forest to forest state) was detected. Each cell in this dataset was assigned either 1 (gain) or 0 (no gain) value.

To assess forest loss or gain, the catchment boundary associated with each stream gauge was used to extract the number of one arc-second pixels that experienced loss (f_l) and gain (f_g) in forest cover over 2000–2012. The percentage of each catchment area that has undergone forest change (f_c) was then derived from the extracted values of each catchment as shown in Eq. (1).

$$f_c = \frac{f_g - f_l}{A} * 100\ (\%) \quad (1)$$

where A is the total number of one arc-second size cells of the catchment.

Resulting from this calculation, f_c can be positive (indicating that the catchment experienced an expansion in forest cover area during 2000–2012), negative (indicating that the catchment experienced a reduction in forest cover area during 2000–2012) or zero (indicating that no change has occurred during 2000–2012). The R-package gfcanalysis (Zvoleff, 2015) was used to download and analyse this data.

2.4. Statistical techniques to assess changes to annual maximum streamflow records

To test for monotonic changes in the annual maximum streamflow time series, a Mann-Kendall test was applied separately to each location (Wilks, 2011). In previous studies (as summarised in Table 1), statistically significant trends were typically reported at the 10% two-sided significance level (with the exception of those studies conducted over the United States). To compare this study with existing studies, the null hypothesis is rejected if the two-sided p-value of the test statistic (Kendall's τ) is lower than 0.1. Depending on the value of τ (i.e. positive or negative), we then infer that there is a monotonic increasing/decreasing trend in the annual maximum streamflow over time.

To assess whether the proportion of trends of hydrology time series was significant over a specific group of observations, a field significance resampling procedure (Wilks, 2011) was applied (Ishak et al., 2013; Kiktev et al., 2003; von Storch and Zwiers, 2002; Westra et al., 2013). The general steps of this approach are summarised as:

(1) The reference period, e.g. {1966, 1967, 1968, 1969, ..., 2005}, was resampled with replacement to create a new set with the same length but different year-order from the original period, for example {2005, 1966, 2001, 1998, ..., 1966}.

(2) The observations at each site were rearranged corresponding to the resampled set of years obtained from step (1) to obtain a new sample of annual maximum streamflow, where any temporal structure was removed but the spatial dependence was preserved.

(3) The Mann-Kendall test was applied to the resampled streamflow at the 10% two-sided significance level, and the percentage of stations showing significant (both increasing and decreasing) trends was recorded.

(4) Steps (1) to (3) were repeated 1000 times to obtain the percentage distribution of stations showing significant trends, representing the distribution of the null hypothesis.

The above approach is appropriate when the data are serially independent; however in this case stations exhibited mild levels of autocorrelation, with a global median value of 0.073, 0.068, and 0.078 for datasets A1, A2 and A3. To account for the effect of autocorrelation on the field significance results across different groups of stations, a moving-blocks bootstrap approach was adopted (Alexander et al., 2006; Kiktev et al., 2007). The distinction

of the moving-blocks bootstrap in comparison to the conventional bootstrap lies in the first step of the resampling procedure (step 1, above), when consecutive years of length L are resampled in order to build up a synthetic sample with a similar level of autocorrelation to the original data. As autocorrelation in the time series varies across different sites, the block length is recommended to be determined based on individual time series (Politis, 2003), and thus L also varies across different sites. To identify optimal block length L for individual annual maximum streamflow time series, an automatic block-length selection procedure (Politis and White, 2004) was implemented using the R script available at http://www.math.ucsd.edu/~politis/SOFT/PPW/ppw.R.

The moving-blocks bootstrap resampling procedure that incorporated the automatic block-length selection process was used for field significance analysis across the experiments in this study and is referred to as the "moving-blocks bootstrap procedure" throughout this paper. Percentages calculated from the observed datasets were compared to resampled distributions generated from the moving-blocks bootstrap procedure to determine whether the annual maximum data of a specific group of stations exhibited monotonic trends. The 10% two-sided significance level of the Mann-Kendall test implies that about 5% of stations would show significant increasing/decreasing trends by random chance. The null hypothesis of no change is rejected when the observed percentage lies outside the 90% confidence interval of the resampled distributions.

Finally, an alternative method to assess the role of catchment characteristics on the presence or absence of trends is to assess whether various attributes of catchments with statistically significant increasing and decreasing trends are similar or not. To do this, a t-test at the 5% significance level was used. If the p-value of a specific test is lower than 0.05, the null hypothesis is rejected and there is evidence to support the statement that there is a significant difference in the attribute subjected to the t test between stations showing increasing trends and stations showing decreasing trends.

2.5. Overall study approach

Based on the above datasets and statistical testing approaches, four main experiments were developed to detect changes in annual maximum streamflow over different spatial domains, and to link any detected changes with climatic and catchment characteristics (see Table 3). The purpose of each experiment, the required datasets and the statistical tests used, is summarised as follow:

– **Identify the global patterns of change in annual maximum streamflow and the influence of reference periods.** Three datasets (A1, A2, and A3) were used to assess changes in the annual maximum flow series and the response of detected trends to changing reference periods. In this experiment, the Mann-Kendall test was applied to each location and the percentage of stations showing significant trends was recorded for each dataset. These numbers were analysed to identify the importance of choosing a reference period in analysing trends of floods at the global scale.
– **Identify the importance of dams to changes in annual maximum streamflow.** In this experiment, the A1 dataset was divided into two subsets (B1 "no dams" and B2 "dams"). The Mann-Kendall test was applied at each location and a bootstrap procedure was applied to create the null hypothesis distribution for each dataset. The comparison between observed percentages of stations showing significant trends and the null hypothesis distribution (obtained from the moving-blocks bootstrap approach) across the three different datasets was used to assess whether the presence of dams has a significant influence on the results of the analysis.
– **Identify the relationship between geographical or climatic characteristics and detected trends.** Dataset B1 ("no dams") was analysed using the Mann-Kendall test and moving-blocks bootstrap procedure. To contrast the relationship between geographical or climatic characteristics and the detected trend, no dam stations were divided into six continents (Africa, Asia, Australia (and some islands), Europe, North America, South America) or nine climatic regions (arid climate, equatorial climate, polar climate, warm summer dry climate, warm winter dry climate, warm fully humid, snow summer dry climate, snow winter dry climate, snow fully humid). Each group was subjected separately to the same statistical analysis to identify differences in significant changes of annual maximum streamflow records across different continents and climatic groups.
– **Assess the effect of catchment area and forest cover changes on trends in annual maximum streamflow.** These two characteristics were chosen as they can affect the amount of water available for runoff. To assess changes to the relationship between flood and catchment characteristics, two separate experiments were conducted for each dataset:
  (1) **Field significance analysis for stations with different catchment characteristics.** In this experiment, all stations were divided into five groups with similar forest cover

Table 3
Experiments conducted, and their associated datasets.

Experiment	Datasets required	Statistical tests	Results
Global pattern of changes in annual maximum streamflow	- Three different streamflow datasets: A1, A2, A3 (see Section 2.3.1)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure)	Section 3.1
Impacts of dam to changes in annual maximum streamflow	- GRanD database was used to classify A1 dataset into B1 (no-dam stations) and B2 (dam stations) datasets - B1 (no dam stations) and B2 (dam stations)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure)	Section 3.2
Relationships between changes in annual maximum streamflow and geographical/climatic characteristics	- Köppen-Weiger climate classification system was used to assign stations into different climatic groups - Dataset B1 (no dam stations)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure)	Section 3.3
Relationships between changes in annual maximum streamflow and catchment characteristics	- Catchment size and proportion of forest cover changes (2000–2014) were used to assign stations into different catchment characteristics group - Dataset B1 (no dam stations)	- Mann-Kendall test - Field significance analysis (moving-blocks bootstrap procedure) - t test	Section 3.4 and Section 3.5

change/catchment sizes using the 20, 40, 60 and 80 quantiles of the frequency distribution. Field significance analysis was applied to these groups to identify whether a group exhibits a significant percentage of stations showing significant trends.

(2) **Apply a *t* test for differences between stations showing increasing and decreasing trends.** Here all stations showing significant trends were divided into two groups: a group of stations with increasing trends and a group with decreasing trends. A *t* test at the 5% significance level was conducted to assess whether the catchment characteristics of increasing trend and decreasing trend groups were indistinguishable. If the *p*-value is lower than 0.05, the null hypothesis is rejected and there is evidence to support that the catchment characteristics between these two groups are different. Based on the distribution of catchment characteristics for these two groups, relationships of these variables to changes in annual maximum streamflow records can be identified.

3. Results

3.1. Patterns of change to global annual maximum streamflow and the influence of different reference periods

We start by investigating the number of stations with statistically significant increasing and decreasing trends at the 10% (two-sided) significance level, and compare this to what would be expected under the null hypothesis (obtained from the moving-blocks bootstrap procedure). The percentage of stations with significant trends is presented in Fig. 4 for each of the three reference periods, with the null hypothesis distribution generated from 1000 conventional bootstrap iterations given as a histogram. As can be seen, over the main reference period (dataset A1; 1966–2005), there were 7.1% of stations with statistically significant increasing trends (corresponding to 136 stations; Fig. 4a), and 11.9% of stations with statistically significant decreasing trends (corresponding to 226 stations; Fig. 4b). The percentage of stations exhibiting statistically significant increasing trends is consistent with the null hypothesis of no change on average across the global dataset, whereas the percentage of stations showing significant decreasing trends is inconsistent with the null hypothesis.

Comparing the results from the main reference period (dataset A1) to the other two reference periods, it is evident that the percentage of stations with significant increasing trends changes from 7.1% (Fig. 4a), to 11.8% for dataset A2 (Fig. 4c) and 16.4% for dataset A3 (Fig. 4e). A similar pattern is evident for stations with decreasing trends, with the percentage of stations showing decreasing trends changing from 11.9% (Fig. 4b), to 15.4% for dataset A2 (Fig. 4d) and 22.9% for dataset A3 (Fig. 4f). With the exception of increasing trends in dataset A1, all results are field significant at the 10% (two-sided) significance level, even when the moving-blocks bootstrap (block size of two) procedure was adopted. Interestingly, more stations showed statistically significant decreasing trends than statistically significant increasing trends for all datasets.

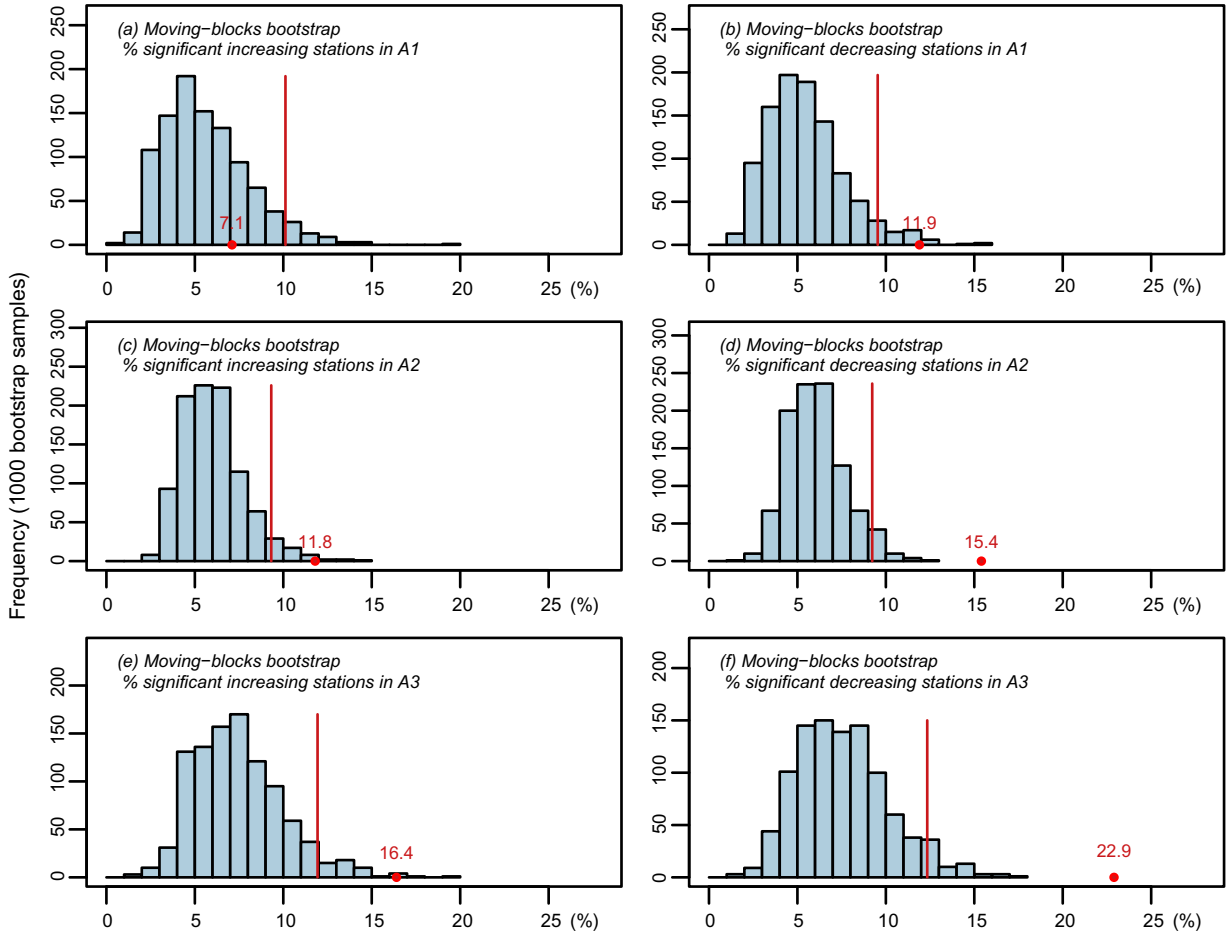


Fig. 4. Percentage of stations showing significant trends based on the Mann–Kendall test with each row displaying increasing and decreasing trends respectively. For dataset A1 (a and b); dataset A2 (c and d); and dataset A3 (e and f). The histogram represents the distribution of significant percentage changes obtained from 1000 iterations of a moving-blocks bootstrap procedure, and the red dot indicates the observed value while the red line indicates the 95th percentile.

One possible explanation for the different results between datasets is the effect of different data lengths on the statistical power of the Mann-Kendall tests, with datasets A1, A2 and A3 having an average of 39.7, 47.6 and 93.0 years of data, respectively (see Section 2.2). For example, of the 652 stations that were common to datasets A1 and A3, there were 105 stations (16.1% of the sample) exhibiting significant changes (either increases or decreases) for dataset A1 compared to 257 stations (39.4% of the sample)

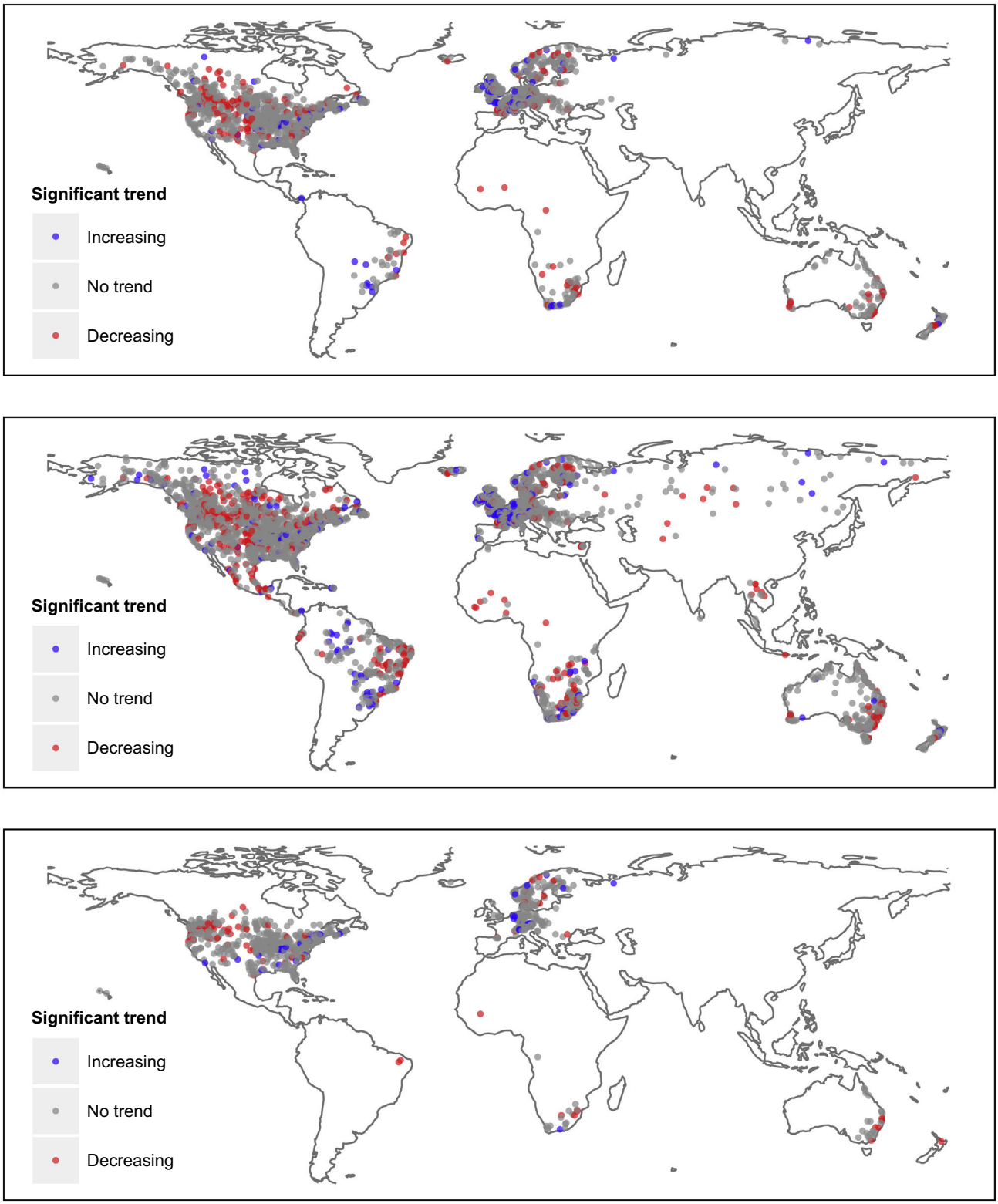


Fig. 5. Results for trends in magnitude of floods events for dataset A1 (upper panel), A2 (middle panel) and A3 (lower panel). Blue (red) filled dots indicate stations with statistically significant increasing (decreasing) trends at the 10% level. Grey dots indicate time series that did not exhibit statistically significant changes at the 10% level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

exhibiting significant changes for dataset A3, indicating a greater likelihood of trend detection for the longer dataset. However, the composition of the stations that were statistically significant was not consistent across the group: in particular, considering again those stations that were common between datasets A1 and A3, approximately half of the stations that exhibited significant trends in group A1 were no longer significant in group A3. Similar patterns were found when comparing common stations between datasets A1 and A2, confirming the potentially strong role of sampling variability combined with potential systematic changes over time at individual gauges. This result is consistent with other studies that find a strong influence of the reference period window on the results of trend detection analyses (Hall et al., 2014; Hannaford et al., 2013; Kundzewicz and Mondiale, 2000; Lins and Slack, 1999; Merz et al., 2012a).

An alternative explanation of the differences between datasets is the potential for shifts in the spatial distribution of the trend results. Spatial plots of stations showing statistically significant increasing and decreasing trends are presented in Fig. 5. Starting with dataset A1 (upper panel of Fig. 5), some regional clustering of the trending stations was evident, with a large number of stations with significant decreasing trends in North America (particularly to the west of the continent), and regions with increasing trends concentrated in parts of Europe (e.g. the UK, France and Germany) and parts of eastern North America.

When these results are compared to the other datasets (middle and lower panels of Fig. 5), it is apparent that although data coverage varies significantly between the datasets, there are important consistencies in the spatial patterns of the trends. For example, the western part of North America and the data-covered regions of Australia (particularly the east coast of the continent) showed a number of stations with decreasing trends, regardless of the dataset used. Similarly, regions with increasing trends—for example in Europe, eastern North America, parts of South America and southern Africa—were generally consistent across the datasets (although data from South America and southern Africa were not available for dataset A3). With a better spatial coverage, dataset A2 also revealed information for regions not adequately covered by datasets A1 and A3: in particular, the number of stations showing increasing trends was outnumbered by the stations showing decreasing trends over Asia with the exception of the far eastern region of Russia, and all stations located in western Africa showed significant decreasing trends. However, due to data limitations over most countries in these regions, and the generous data admission criteria for dataset A2 (allowing up to 50% of missing data), these results should be interpreted with care.

Based on the above analyses, it can be concluded that although individual stations exhibiting statistically significant trends were often different between reference periods, the above spatial analysis indicates some consistency between the reference periods at the regional scale. Furthermore, stations showing significant increasing trends were outnumbered by stations showing significant decreasing trends in all three datasets, indicating that although there may be evidence of regional increasing trends in

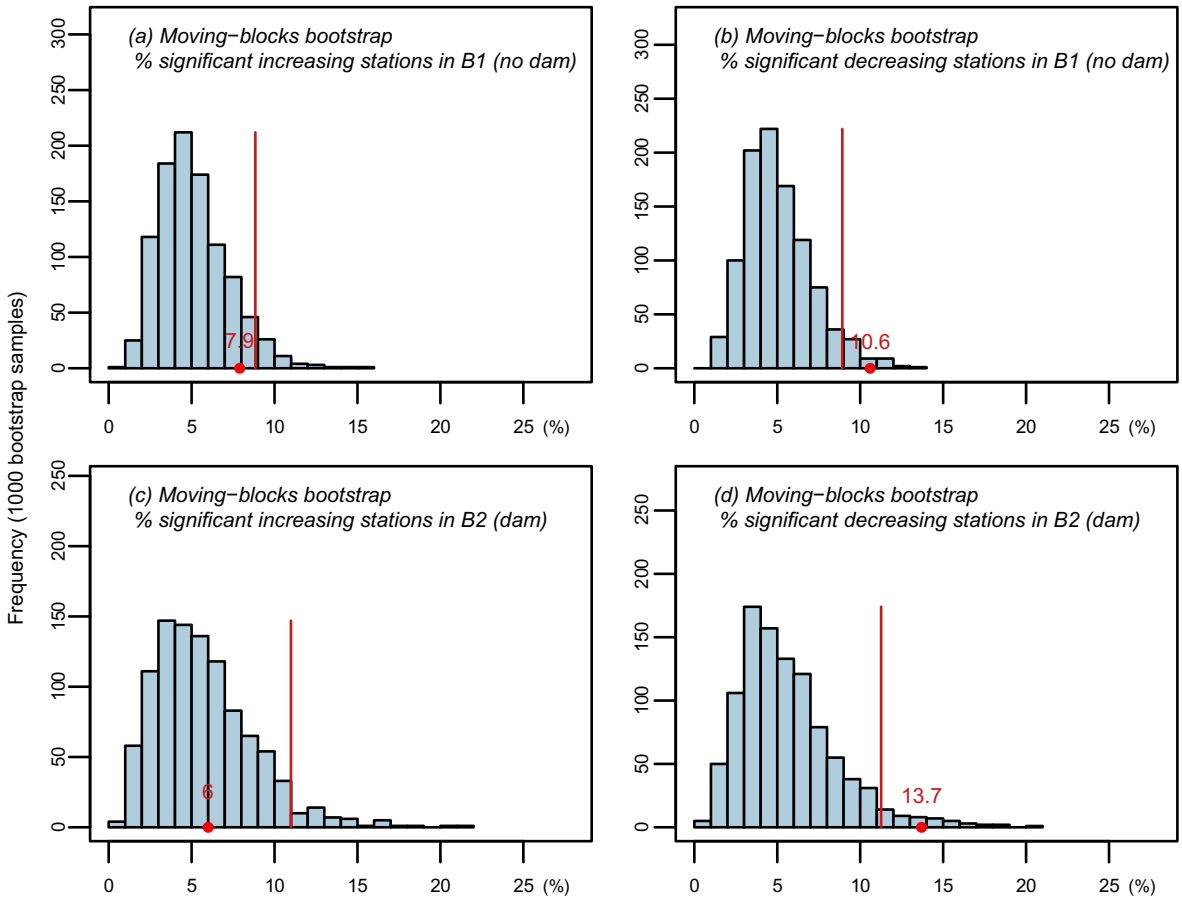


Fig. 6. Percentage of stations showing significant trends based on the Mann–Kendall test for 1143 stations in "no dams" (B1) dataset (a and b), and 746 stations in "dams" (B2) dataset (c and d). Left panels show the results for percentage of stations showing significant increasing trend while right panels show results for percentage of stations showing significant decreasing trend. The histogram represents the distribution of percentages obtained from 1000 moving-blocks bootstrap iterations; the red dot indicates the observed value while the red line indicates the 95th percentile.

flood hazard, the hypothesis that there is a significant increase in flood hazard when averaged over all the data-covered regions of the globe is not supported by this analysis.

There are multiple systematic factors that could explain these findings, which are investigated in the following sections, such as the influence of dams and other forms of human activity. As dataset A1 provides a reasonable compromise between data length and spatial coverage, we focus on this dataset in the following sections.

3.2. Impact of dams on the number of stations showing significant changes and the importance of resampling technique

To assess impacts of dams on trends in annual maximum streamflow records, field significance analyses using moving-blocks bootstrap were applied separately to the "no dams" group (B1) and "dams" group (B2), and the results compared with the full reference period (A1). The results are given in Fig. 6, and show that 46 stations (6.0%) had statistically increasing trends in the "dams" group compared to 90 stations (7.9%) for the "no dams" group, whereas 105 stations (13.7%) had statistically significant decreasing trends for the "dams" group compared to 121 stations (10.6%) for the "no dams" group. The percentage of stations showing statistically significant increasing trends was not field significant (Fig. 6a and e), whereas the percentage of stations showing statistically significant decreasing trends was field significant (Fig. 6c and g) when either B1 or B2 was used. The results were therefore generally consistent with the reference dataset A1, indicating that the effect of dams does not appear to have a substantive influence on the overall trend results.

Although the above results suggest that the presence of large dams does not substantially influence the overall trend results, from a hydrological perspective, large dams would be expected to have a significant effect on flood flows, as in many cases the dams are designed to reduce flood magnitude and the flood damage on human assets (FitzHugh and Vogel, 2011; Jaramillo and Destouni, 2015; Lajoie et al., 2007). On this principle, the remaining analyses are focused on the "no dams" case (i.e. dataset B2) to minimise the influence of large hydraulic structures on any trend results.

3.3. Relationships between trends in floods and geographical/climatic characteristics

This section evaluates whether trends can be associated with the geographical or climatic characteristics of stations. To do this, all stations were divided into either six continents (representing similar geographical characteristics) or nine Köppen climate zones (representing similar climatic characteristics). The results of the continent-based analysis are shown in Table 4, with three continents (North America, Australia and Africa) showing more stations with decreasing trends than increasing trends, and three continents (South America, Asia and Europe) showing more stations with increasing trends than decreasing trends. However, as with the results shown in Fig. 5, the data were not evenly distributed between continents. For example, the whole of Asia is covered by only two streamflow locations, and thus is not sufficient to constitute a representative sample.

The results of the field significance analysis using the moving-blocks bootstrap technique for each continent (which accounts for the number of stations in the record) are consistent with what was found at the global scale (Section 3.1), with the percentages of stations showing significant decreasing trends being field

Table 4
Summary of Mann-Kendall test and field significance analysis across different continents.

	Asia	North America	Europe	South America	Australia	Africa	Global
Available stations	2	418	560	26	79	58	1143
Number of increasing trend stations	1 (50.0%)	11 (2.6%)	67 (*12.0%)	6 (*23.1%)	2 (2.5%)	3 (5.2%)	90 (7.9%)
Number of decreasing trend stations	0 (0%)	55 (*13.2%)	37 (6.6%)	5 (*19.2%)	15 (*19.0%)	9 (15.5%)	121 (*10.6%)

(*: field significant at 10% level).

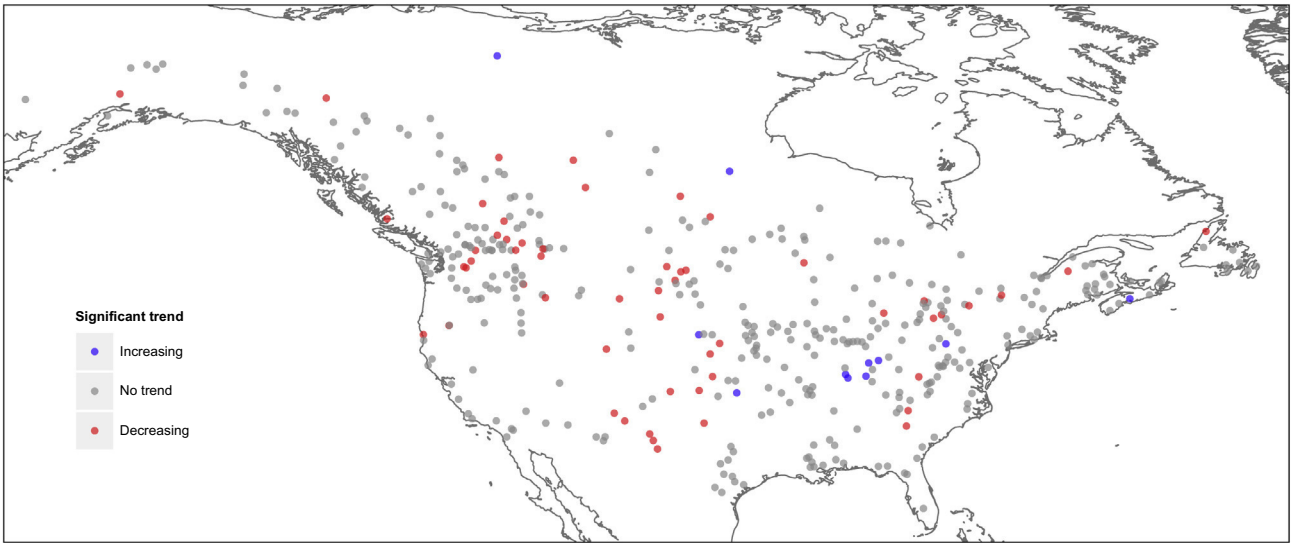


Fig. 7. Trends in magnitude of annual maximum flow events for stations in the B1 dataset over North America. Blue (red) filled dots indicate stations with a statistically significant increasing (decreasing) trend at the 10% level. Grey dots indicate the time series that did not exhibit statistically significant changes at the 10% level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significant in three continents (13.2%, 19.2% and 19.0% for North America, South America and Australia respectively), whereas Europe and South America show statistical evidence of increasing trends (12.0% and 23.1% respectively).

Focusing on continents well-represented in the database (i.e. Europe and North America), the most notable trends are in the western part of North America (as shown in Fig. 7), where there are approximately ten times as many stations with significant decreasing trends compared to significant increasing trends. This was also shown in previous studies of the United States (Lins and Slack, 1999), and Canada (Cunderlik and Ouarda, 2009; Westmacott and Burn, 1997; Whitfield, 2001; Zhang et al., 2001), which generally have found a decrease in the magnitude of floods over time that might be at least partially attributable to changes in snowmelt timing. In other parts of North America and Europe, stations with significant increasing or decreasing trends appear to be distributed more randomly and this is consistent with previous studies of the United States (Lins and Slack, 1999; Mallakpour and Villarini, 2015; Villarini et al., 2009) and Europe (Hannaford et al., 2013; Stahl et al., 2011).

In addition to grouping the results by continent, stations are also grouped by climatic zone (Table 5), and the results provide further consistent evidence to what have been found at global scale. As can be seen, field significant decreasing trends are observed with the moving-blocks bootstrap technique in: equatorial climates, arid climates, and snow fully humid climates. In contrast, warm fully humid is the only climate zone shows field significant increasing trends. Table 5 also demonstrates the inhomogeneity in data coverage across climatic regions. In particular, more than 83% of stations are located in the climatic regions "warm fully humid" and "snow fully humid", corresponding to the large proportion of these regions in North America and Europe.

3.4. Relationships between trends in floods and catchment area

Catchment area has been used in a number of studies as an indicator of the magnitude of land-use change (Hannaford et al., 2013; Ishak et al., 2013; Stahl et al., 2010), and we now explore the influence of area on the overall trend results. To identify the relationship between trends in floods and catchment area, the 1143 "no dam" stations were divided into five groups with an equal number of samples based on their catchment area. Similar to previous sections, field significance was assessed using moving-blocks bootstrap procedure. Of all 1143 stations, catchment area varies from 1 km^2 to 500,000 km^2, with more than 60% of the catchments having areas less than ~1400 km^2. The 20, 40, 60 and 80 quantiles of the catchment area distribution were used as thresholds to create five different groups of stations based on their area: "very small", "small", "medium", "large" and "very large" catchments. As shown in Table 6, the percentage of stations with decreasing trends was found to be significant for the three latter groups, while the percentage of stations exhibiting statistically significant increasing trends was field significant for only "very small" group.

A notable pattern in the percentages of stations showing significant trends over these catchment-size groups is that groups with larger catchment areas generally have stations with more (less) significant decreasing (increasing) trends. As illustrated in Fig. 8, the percentage of stations showing significant decreasing trends is three times the percentage of stations showing significant increasing trends in the last two groups (catchments with area above 1368 km^2). Even in stations in the "medium catchment" group, stations showing significant decreasing trends also outnumbered those showing significant increasing trends. In contrast, the "very small" and "small" catchment groups generally have more stations showing significant increasing trends than stations showing decreasing trends. An implication of this pattern is that the strategy of choosing stations using catchment area may have a substantial influence on the detected changes in annual maximum streamflow records. For example, analysing changes in annual maximum streamflow records using only small catchments (e.g. catchments with area less than 500 km^2) will potentially lead to results where there are more stations showing significant increasing trends.

As an alternative approach to analysing the impact of catchment size on trend results, all stations showing significant trends were divided into stations showing significant increasing trends and stations showing significant decreasing trends, to determine

Table 5
Percentage of stations showing significant increasing (decreasing) trend at the 10% significance level over climatic regions.

Climate groups	No. of stations	Number of increasing trend stations	Number of decreasing trend stations
Equatorial climates (A)	24	3 (12.5%)	5 (*20.8%)
Arid climates (B)	52	2 (3.8%)	19 (*36.5%)
Warm summer dry (Cs)	50	1 (2.0%)	7 (14.0%)
Warm winter dry (Cw)	22	0 (0%)	3 (13.6%)
Warm fully humid (Cf)	562	60 (*10.7%)	38 (6.8%)
Snow summer dry (Ds)	20	0 (0%)	2 (10.0%)
Snow winter dry (Dw)	1	0 (0%)	0 (0%)
Snow fully humid (Df)	386	20 (5.2%)	45 (*11.7%)
Polar climates (E)	26	4 (15.4%)	2 (7.7%)
Total	1143	90 (7.9%)	121 (*10.6%)

(*: field significant at 10% level).

Table 6
Percentage of stations showing significant increasing (decreasing) trends at the 10% significance level over different catchment-size groups.

	Catchment size groups				
	Very small catchments	Small catchments	Medium catchments	Large catchments	Very large catchments
Number of stations	229	228	229	228	229
Significant increasing	26 (*11.4%)	22 (9.6%)	21 (9.2%)	10 (4.4%)	11 (4.8%)
Significant decreasing	22 (9.6%)	11 (4.8%)	28 (*12.2%)	27 (*11.8%)	33 (*14.4%)
Size of catchment (km^2)	1–134	134–389	389–1368	1368–4008	4008–500,000

(*: field significant at 10% level).

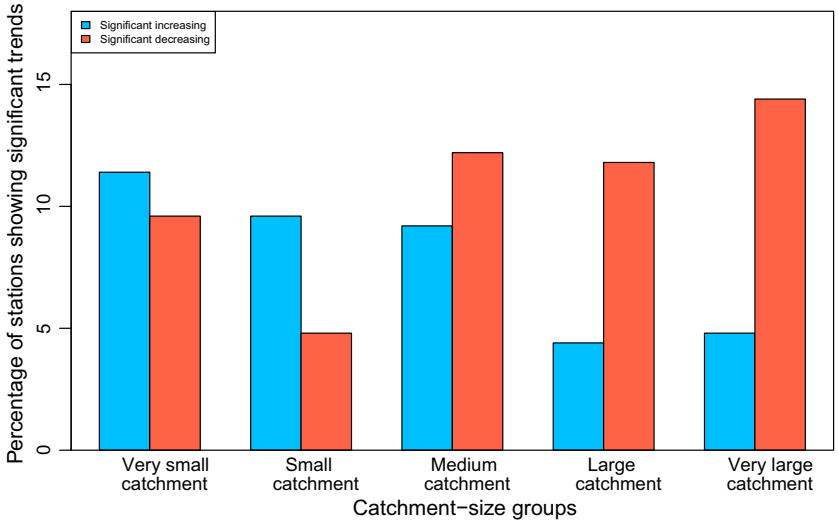


Fig. 8. Bar chart of the differences in area between stations showing significant increasing and decreasing trends across five different catchment size groups.

Table 7
Percentage of stations showing significant increasing (decreasing) trend at the 10% significance level over forest cover change groups.

	Forest changes groups		
	Forest loss ($f_c < 0$)	No change ($f_c = 0$)	Forest gain ($f_c > 0$)
Number of stations	277 (24%)	89 (8%)	777 (68%)
Stations showing increasing trends	20 (7.2%)	10 (11.2%)	60 (7.7%)
Stations showing decreasing trends	32 (*11.6%)	9 (10.1%)	80 (*10.3%)

(*: field significant at 10% level).

if the catchment areas between these two groups are different from each other. Interestingly, the results of the *t* test for catchments did not provide evidence to reject the hypothesis that there is no significant difference in catchment area between the two groups. Although the mean of catchment size of the significant decreasing trend group is higher (9932 km^2 in comparison with 4868 km^2 in the significant increasing trend group), the null hypothesis is not rejected (*p*-value: 0.32). This indicates that although there is a clear pattern in trends as indicated by Fig. 8, a high level of random variability is also present and the specific nature of the analyses is likely to have a large effect on the interpretation of the results.

3.5. Relationships between trends in floods and changes in forest cover rate

Finally, we investigate the role of forest cover on any identified trends in annual maximum streamflow, focusing again on the 1143 "no dam" catchments. Each station was assigned into one among three different groups based on the percentage of changes in catchment area over 2000–2012: forest gain ($f_c > 0$), no change ($f_c = 0$) and forest loss ($f_c < 0$), with results shown in Table 7. Interestingly, 68% of the catchments have experienced an expansion in forest cover, whereas only 24% of catchments have experienced a decline in forest cover, with an average proportion of total forest area cover change across all catchments being 1.1% over the 2000–2012 period. These numbers are slightly different to changes in forest area globally as reported in the recent Global Forest Resources Assessment (Keenan et al., 2015), which showed a declining proportion of 1.4% of global forest area over the same period. This may be caused by the inhomogeneous spatial coverage of the GRDC data, with the Keenan et al. (2015) analysis conducted on all continental areas except for Antarctica, whereas our results are based only on those regions with streamflow data.

To assess whether the number of stations showing significant changes is field significant, the moving-blocks bootstrap procedure was applied to each group separately. Firstly, considering the stations with increasing trends, the "forest loss" and "forest gain" datasets show a similar magnitude of increasing trends compared to the global scale "no dam" analysis (Section 3.2). In contrast, the dataset with "no change" in forest cover showed a higher percentage change, but this result is not statistically significant and may be due to the relatively small sample size of this dataset. Considering next the decreasing trends, the percentage changes were consistent with the global scale "no dam" analysis for both "forest loss" and "forest gain" as field significance was detected.

As an alternative approach to testing the role of forest cover changes, the data were divided into quintiles by forest cover change (i.e. dividing the data at the 20, 40, 60 and 80 percentiles), and the field significance of these results were evaluated. The results (not shown) were consistent with the results presented in Table 7.

4. Discussion and conclusions

Whilst acknowledging data concerns regarding data record length, spatial coverage and quality, this study provides an indication of changes to annual maximum streamflow in the data-covered regions of the globe. Three reference datasets representing different compromises between data length and spatial coverage were examined, and it was found that the choice of dataset had a significant influence in terms of whether individual stations had significant increasing or decreasing trends, indicating a large role of sampling variability and other factors on the station-level trend

results. However, when comparing trend results of the three datasets at the global scale, it was found that a greater number of stations exhibited statistically significant trends (both increasing and decreasing) as the average length of record increased, with almost 40% of stations exhibiting statistically significant trends at the 10% significance level for the longest dataset (A3). This is at least partially explained by the increased statistical power of trend detection methods as the record becomes longer.

Although there were substantial differences in station-level results depending on the dataset used, the results were more consistent when viewed at the regional scale. In particular, decreasing trends were observed for a large number of stations in western North America and the data-covered regions of Australia, and increasing trends in parts of Europe, eastern North America, parts of South America and southern Africa. These changes were also consistent with previous regional studies; for instance, previous studies have found decreasing trends in the western part of North America (Cunderlik and Ouarda, 2009; Lins and Slack, 1999; Westmacott and Burn, 1997; Whitfield, 2001; Zhang et al., 2001) and in Australia (Ishak et al., 2013), while the eastern part of North America and Europe generally have a mixed pattern of changes (Hannaford et al., 2013; Lins and Slack, 1999; Mallakpour and Villarini, 2015; Stahl et al., 2011; Villarini et al., 2009). In addition to dividing the result by geographical location, consistent results were also found when the data were broken into separate climate groups, with decreasing trends particularly apparent in stations with equatorial, arid and snow fully humid climates (field significant at the 10% significance level) while increasing trends are present in only stations with warm fully humid climates.

Neither the presence of dams nor changes in forest cover had a large effect on the trend results (although most analyses were conducted using the "no dam" dataset to minimise the potential influence of large hydraulic structures on results). The lack of a conclusive finding regarding the influence of forest cover suggests that forest cover may not play a dominant role explaining trends in flood hazard (Bradshaw et al., 2007; Bruijnzeel, 2004; Clark, 1987; Van Dijk et al., 2009), although it should be cautioned that the results could also be explained by other confounding factors (e.g. agricultural development, urbanisation) that may co-occur alongside changes in forest cover. It should also be noted that although the Global Forest Change dataset is the most advanced product currently available of global tree canopy, the period of this dataset does not fully overlap the streamflow records, which is a major limitation of this analysis.

Importantly, for almost all datasets considered, and regardless of whether the stations were filtered by the presence of dams, catchment area, or forest cover changes, there were more stations exhibiting significant decreasing trends than increasing trends. The results of our global analysis were therefore generally consistent with Kundzewicz et al. (2004), who also showed that the number of stations showing significant increasing trends was found to be smaller than the number of stations showing significant decreasing trends. The exception of our analysis was for the "no dam" dataset for the smallest two quintiles by catchment size, where there were more stations exhibiting increasing than decreasing trends. This result was inconsistent with what was found in Milly et al. (2002), although that study focused only on the frequency of extreme events (i.e. discharges exceeding 100-year levels), whereas our study focused on annual maximum streamflow and thus includes relatively moderate 'flood' events as well as rarer events. There are multiple factors that might explain the effect of catchment size on the trend results; for instance, larger catchments typically are affected by longer time-scale rainfall events, and thus the climatic forcings may be different compared to smaller catchments (Westra et al., 2014). Furthermore, catchment size may be associated with other anthropogenic modifications in the catchment, such as the presence of agriculture, urbanisation, and the construction of a range of hydraulic infrastructure not included in the large dams database.

Despite potential concerns about data quality, one interesting pattern to emerge was that detected changes in annual maximum streamflow are inconsistent with the evidence of trends in precipitation. At the global scale, annual maximum precipitation intensities were found to have increased (Min et al., 2011) and a large-scale increasing pattern in extreme precipitation was detected (Lehmann et al., 2015), with North America experiencing more increasing trends than decreasing trends in annual maximum precipitation (Westra et al., 2013). These precipitation-based results therefore appear to be inconsistent with the trend results for floods found in our analysis, and indicate the potentially important role of changes in catchment conditions and river morphology to changes in streamflow regimes (Hall et al., 2014; Merz et al., 2012a). Further research is needed to quantify the contribution of catchment condition to the rainfall-runoff relationship at global and regional scales, including investigation of changes in other dimensions of flooding, such as their duration, volume, and intensity.

Finally, the changes in the flood hazard as assessed in this study do not explain observed increases in flood losses (Kundzewicz et al., 2013; Mills, 2005) or in the number of reported events (Munich Re, 2015; Swiss Re, 2015). In our results, there were more decreasing trends than increasing trends in almost all cases, regardless of the dataset used, whether catchments were filtered by climatic condition, presence of dams, or forest cover loss. This indicates that other factors contributing to flood risks (i.e. exposure and vulnerability), are likely to contribute a higher share toward the rise of flood losses at a global scale.

Acknowledgement

We thank the Global Runoff Data Centre (GRDC) for providing the observed streamflow data. Mr Hong Xuan Do receives financial support from the Australia Awards Scholarships (AAS). Dr Westra's time was supported by Australian Research Council Discovery project DP150100411. The authors also wish to thank two anonymous reviewers for their constructive comments and suggestions.

References

Alexander, L. et al., 2006. Global observed changes in daily climate extremes of temperature and precipitation. J. Geophys. Res.: Atmos. 111 (D5).

Archfield, S.A. et al., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. 51 (12), 10078–10091. http://dx.doi.org/10.1002/2015WR017498.

Archfield, S.A., Hirsch, R.M., Viglione, A., Blöschl, G., 2016. Fragmented patterns of flood change across the United States. Geophys. Res. Lett. http://dx.doi.org/10.1002/2016GL070590. n/a-n/a.

Arnell, N., Gosling, S., 2014. The impacts of climate change on river flood risk at the global scale. Climatic Change 1–15. http://dx.doi.org/10.1007/s10584-014-1084-5.

Bard, A., Renard, B., Lang, M., 2012. Floods in the Alpine Areas of Europe, Changes in Flood Risk in Europe. CRC Press. http://dx.doi.org/10.1201/b12348-23, pp. 362–371.

Bierkens, M.F.P., 2015. Global hydrology 2015: state, trends, and directions. Water Resour. Res. 51 (7), 4923–4947. http://dx.doi.org/10.1002/2015WR017173.

Bradshaw, C.J.A., Sodhi, N.S., Peh, K.S.H., Brook, B.W., 2007. Global evidence that deforestation amplifies flood risk and severity in the developing world. Glob. Change Biol. 13 (11), 2379–2395. http://dx.doi.org/10.1111/j.1365-2486.2007.01446.x.

Bruijnzeel, L.A., 2004. Hydrological functions of tropical forests: not seeing the soil for the trees? Agric. Ecosyst. Environ. 104 (1), 185–228.

Burn, D.H. et al., 2012. Reference hydrologic networks II. Using reference hydrologic networks to assess climate-driven changes in streamflow. Hydrol. Sci. J. 57 (8), 1580–1593. http://dx.doi.org/10.1080/02626667.2012.728705.

Burn, D.H., Whitfield, P.H., 2016. Changes in floods and flood regimes in Canada. Can. Water Resour. J./Revue canadienne des ressources hydriques 41 (1–2), 139–150. http://dx.doi.org/10.1080/07011784.2015.1026844.

Buschman, F.A., Hoitink, A.J.F., van der Vegt, M., Hoekstra, P., 2009. Subtidal water level variation controlled by river flow and tides. Water Resour. Res. 45 (10). http://dx.doi.org/10.1029/2009WR008167.

Clark, C., 1987. Deforestation and floods. Environ. Conserv. 14 (01), 67–69.
Clark, M.P. et al., 2015. Improving the representation of hydrologic processes in Earth System Models. Water Resour. Res. 51 (8), 5929–5956. http://dx.doi.org/10.1002/2015WR017096.
CRED, 2015. The Human Cost of Natural Disasters: A Global Perspective. Centre for Research on the Epidemiology of Disasters, Brussels.
Cunderlik, J.M., Ouarda, T.B.M.J., 2009. Trends in the timing and magnitude of floods in Canada. J. Hydrol. 375 (3–4), 471–480. http://dx.doi.org/10.1016/j.jhydrol.2009.06.050.
Dankers, R. et al., 2014. First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. Proc. Natl. Acad. Sci. 111 (9), 3257–3261.
Delgado, J., Apel, H., Merz, B., 2010. Flood trends and variability in the Mekong river. Hydrol. Earth Syst. Sci. 14 (3), 407–418.
Di Baldassarre, G., Claps, P., 2011. A hydraulic study on the applicability of flood rating curves. Hydrol. Res. 42 (1), 10–19. http://dx.doi.org/10.2166/nh.2010.098.
Douglas, E., Vogel, R., Kroll, C., 2000. Trends in floods and low flows in the United States: impact of spatial correlation. J. Hydrol. 240 (1), 90–105.
FitzHugh, T.W., Vogel, R.M., 2011. The impact of dams on flood flows in the United States. River Res. Appl. 27 (10), 1192–1215.
Franks, S.W., Kuczera, G., 2002. Flood frequency analysis: evidence and implications of secular climate variability, New South Wales. Water Resour. Res. 38 (5), 20-1–20-7. http://dx.doi.org/10.1029/2001WR000232.
Fuchs, T., Schneider, U., Rudolf, B., 2007. Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC). Deutscher Wetterdienst, Offenbach.
Ghasemizade, M., Schirmer, M., 2013. Subsurface flow contribution in the hydrological cycle: lessons learned and challenges ahead—a review. Environ. Earth Sci. 69 (2), 707–718. http://dx.doi.org/10.1007/s12665-013-2329-8.
GRDC, 2015. Report of the Twelfth Meeting of the GRDC Steering Committee, Koblenz, Germany, 18–19 June 2014, Global Runoff Data Centre (GRDC), Koblenz, Germany. doi: 10.5675/GRDC_Report_46.
Guha-Sapir, D., Hoyois, P., Below, R., 2015. Annual Disaster Statistical Review 2014: The numbers and trends, UCL.
Guha-Sapir, D., Hoyois, P., Below, R., 2014. Annual Disaster Statistical Review 2013: The Numbers and Trends. CRED, Brussels.
Hall, J. et al., 2014. Understanding flood regime changes in Europe: a state-of-the-art assessment. Hydrol. Earth Syst. Sci. 18 (7), 2735–2772. http://dx.doi.org/10.5194/hess-18-2735-2014.
Hannaford, J., Buys, G., Stahl, K., Tallaksen, L.M., 2013. The influence of decadal-scale variability on trends in long European streamflow records. Hydrol. Earth Syst. Sci. 17 (7), 2717–2733. http://dx.doi.org/10.5194/hess-17-2717-2013.
Hannaford, J., Marsh, T., 2006. An assessment of trends in UK runoff and low flows using a network of undisturbed catchments. Int. J. Climatol. 26 (9), 1237–1253. http://dx.doi.org/10.1002/joc.1303.
Hannaford, J., Marsh, T.J., 2008. High-flow and flood trends in a network of undisturbed catchments in the UK. Int. J. Climatol. 28 (10), 1325–1338. http://dx.doi.org/10.1002/joc.1643.
Hannah, D.M. et al., 2011. Large-scale river flow archives: importance, current status and future needs. Hydrol. Process. 25 (7), 1191–1200. http://dx.doi.org/10.1002/hyp.7794.
Hansen, M.C. et al., 2013. High-resolution global maps of 21st-century forest cover change. Science 342 (6160), 850–853. http://dx.doi.org/10.1126/science.1244693.
Herschy, R., 1994. The analysis of uncertainties in the stage-discharge relation. Flow Meas. Instrum. 5 (3), 188–190. http://dx.doi.org/10.1016/0955-5986(94)90018-3.
Hewlett, J.D., Hibbert, A.R., 1967. Factors affecting the response of small watersheds to precipitation in humid areas.
IPCC, 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press, Cambridge, UK, and New York, NY, USA.
IPCC, 2014. Summary for policymakers. In: Field, C.B. et al. (Eds.), Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 32.
Ishak, E., Rahman, A., Westra, S., Sharma, A., Kuczera, G., 2013. Evaluating the non-stationarity of Australian annual maximum flood. J. Hydrol. 494, 134–145.
Ivancic, T., Shaw, S., 2015. Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. Climatic Change 1–13. http://dx.doi.org/10.1007/s10584-015-1476-1.
Jaramillo, F., Destouni, G., 2015. Local flow regulation and irrigation raise global human water consumption and footprint. Science 350 (6265), 1248–1251. http://dx.doi.org/10.1126/science.aad1010.
Johnson, F. et al., 2016. Natural hazards in Australia: floods. Climatic Change 1–15. http://dx.doi.org/10.1007/s10584-016-1689-y.
Keenan, R.J. et al., 2015. Dynamics of global forest area: results from the FAO Global Forest Resources Assessment 2015. For. Ecol. Manage. 352, 9–20. http://dx.doi.org/10.1016/j.foreco.2015.06.014.
Kiktev, D., Caesar, J., Alexander, L.V., Shiogama, H., Collier, M., 2007. Comparison of observed and multimodeled trends in annual extremes of temperature and precipitation. Geophys. Res. Lett. 34 (10).
Kiktev, D., Sexton, D.M., Alexander, L., Folland, C.K., 2003. Comparison of modeled and observed trends in indices of daily climate extremes. J. Clim. 16 (22), 3560–3571.
Kron, W., 2005. Flood risk = hazard • values • vulnerability. Water Int. 30 (1), 58–68. http://dx.doi.org/10.1080/02508060508691837.
Kundzewicz, Z.W. et al., 2004. Detection of Change in World-Wide Hydrological Time Series of Maximum Annual Flow. Global Runoff Date Centre, Koblenz, Germany.
Kundzewicz, Z.W. et al., 2013. Flood risk and climate change: global and regional perspectives. Hydrol. Sci. J. 59 (1), 1–28. http://dx.doi.org/10.1080/02626667.2013.857411.
Kundzewicz, Z.W., Mondiale, O.M., 2000. Detecting trend and other changes in hydrological data. WMO.
Kundzewicz, Z.W. et al., 2012. Changes in Flood Risk–Setting the Stage, Changes in Flood Risk in Europe. CRC Press, 11–26.
Lajoie, F., Assani, A.A., Roy, A.G., Mesfioui, M., 2007. Impacts of dams on monthly flow characteristics. The influence of watershed size and seasons. J. Hydrol. 334 (3), 423–439.
Lehmann, J., Coumou, D., Frieler, K., 2015. Increased record-breaking precipitation events under global warming. Climatic Change 1–15. http://dx.doi.org/10.1007/s10584-015-1434-y.
Lehner, B. et al., 2011. Global reservoir and dam (grand) database. Technical Documentation, Version, 1.
Lehner, B., Verdin, K., Jarvis, A., 2006. HydroSHEDS technical documentation, version 1.0. World Wildlife Fund US, Washington, DC: 1–27.
Lins, H.F., Slack, J.R., 1999. Streamflow trends in the United States. Geophys. Res. Lett. 26 (2), 227–230.
Mallakpour, I., Villarini, G., 2015. The changing nature of flooding across the central United States. Nat. Clim. Change 5 (3), 250–254. http://dx.doi.org/10.1038/nclimate2516.
Merz, B., Kundzewicz, Z., Delgado, J., Hundecha, Y., Kreibich, H., 2012a. Detection and attribution of changes in flood hazard and risk. Changes in flood risk in Europe. IAHS Special Publication 10, 435–458.
Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., Hundecha, Y., 2012b. HESS opinions "More efforts and scientific rigour are needed to attribute trends in flood time series". Hydrol. Earth Syst. Sci. 16 (5), 1379–1387.
Micevski, T., Franks, S.W., Kuczera, G., 2006. Multidecadal variability in coastal eastern Australian flood data. J. Hydrol. 327 (1–2), 219–225. http://dx.doi.org/10.1016/j.jhydrol.2005.11.017.
Mills, E., 2005. Insurance in a climate of change, Science, 309, 1040+.
Milly, P.C.D. et al., 2015. On critiques of "stationarity is dead: whither water management?". Water Resour. Res. 51 (9), 7785–7789. http://dx.doi.org/10.1002/2015WR017408.
Milly, P.C.D., Wetherald, R.T., Dunne, K., Delworth, T.L., 2002. Increasing risk of great floods in a changing climate. Nature 415 (6871), 514–517.
Min, S.-K., Zhang, X., Zwiers, F.W., Hegerl, G.C., 2011. Human contribution to more-intense precipitation extremes. Nature 470 (7334), 378–381. http://dx.doi.org/10.1038/nature09763.
Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int. J. Climatol. 25 (6), 693–712. http://dx.doi.org/10.1002/joc.1181.
Morán-Tejeda, E., López-Moreno, J.I., Vicente-Serrano, S.M., Lorenzo-Lacruz, J., Ceballos-Barbancho, A., 2012. The contrasted evolution of high and low flows and precipitation indices in the Duero basin (Spain). Hydrol. Sci. J. 57 (4), 591–611. http://dx.doi.org/10.1080/02626667.2012.673722.
Munich Re, 2015. NatCatSERVICE: Loss events worldwide 1980–2014 Munich Re, Munich.
Nelson, B., 2009. Data sharing: empty archives. Nature, 461.
Nka, B.N., Oudin, L., Karambiri, H., Paturel, J.E., Ribstein, P., 2015. Trends in floods in West Africa: analysis based on 11 catchments in the region. Hydrol. Earth Syst. Sci. 19 (11), 4707–4719. http://dx.doi.org/10.5194/hess-19-4707-2015.
Peduzzi, P., Dao, H., Herold, C., Mouton, F., 2009. Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index. Nat. Hazards Earth Syst. Sci. 9 (4), 1149–1159. http://dx.doi.org/10.5194/nhess-9-1149-2009.
Politis, D.N., 2003. The impact of bootstrap methods on time series analysis. Stat. Sci. 18 (2), 219–230.
Politis, D.N., White, H., 2004. Automatic block-length selection for the dependent bootstrap. Econometric Rev. 23 (1), 53–70.
Rubel, F., Kottek, M., 2010. Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. Meteorol. Z. 19 (2), 135–141. http://dx.doi.org/10.1127/0941-2948/2010/0430.
Schilling, K.E., Libra, R.D., 2003. Increased baseflow in Iowa over the second half of the 20th century. JAWRA J. Am. Water Resour. Assoc. 39 (4), 851–860. http://dx.doi.org/10.1111/j.1752-1688.2003.tb04410.x.
Sheng, Y., Linghui, W., 2012. Detection of Changes, Changes in Flood Risk in Europe. CRC Press. http://dx.doi.org/10.1201/b12348-26, pp. 387-408.
Slater, L.J., Singer, M.B., Kirchner, J.W., 2015. Hydrologic versus geomorphic drivers of trends in flood hazard. Geophys. Res. Lett. 42 (2), 370–376. http://dx.doi.org/10.1002/2014GL062482.
Slater, L.J., Villarini, G., 2016. Recent trends in U.S. flood risk. Geophys. Res. Lett. 43 (24), 12428–12436. http://dx.doi.org/10.1002/2016GL071199.
Sood, A., Smakhtin, V., 2015. Global hydrological models: a review. Hydrol. Sci. J. 60 (4), 549–565. http://dx.doi.org/10.1080/02626667.2014.950580.
Stahl, K. et al., 2010. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. Hydrol. Earth System Sci., 14, pp. 2367–2382.
Stahl, K., Tallaksen, L., Hannaford, J., Van Lanen, H., 2011. Abflusstrends in Europa: Vergleich eines Multi-Modell-Experiments mit Beobachtungen. Blöschl, G. und Merz, R. (Hrsg.): Hydrologie & Wasserwirtschaft-von der Theorie zur Praxis. Beiträge zum Tag der Hydrologie, 24: 25.

Stover, S.C., Montgomery, D.R., 2001. Channel change and flooding, Skokomish River, Washington. J. Hydrol. 243 (3–4), 272–286. http://dx.doi.org/10.1016/S0022-1694(00)00421-2.

Swiss Re, 2015. Natural catastropes and man-made disaster in 2014, Swiss Reinsurance Company, Zurich, Switzerland.

Turner, M., Bari, M., Amirthanathan, G., Ahmad, Z., 2012. Australian network of hydrologic reference stations-advances in design, development and implementation.

Van Dijk, A.I.J.M. et al., 2009. Forest–flood relation still tenuous – comment on 'Global evidence that deforestation amplifies flood risk and severity in the developing world' by C. J. A. Bradshaw, N.S. Sodi, K. S.-H. Peh and B.W. Brook. Global Change Biology, 15(1), 110–115. doi: 10.1111/j.1365-2486.2008.01708.x.

Villarini, G., Serinaldi, F., Smith, J.A., Krajewski, W.F., 2009. On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resour. Res., 45(8), n/a-n/a. doi: 10.1029/2008WR007645.

von Storch, H., Zwiers, F.W., 2002. Statistical analysis in climate research. Citeseer.

Ward, P.J. et al., 2015. Usefulness and limitations of global flood risk models. Nat. Clim. Change 5 (8), 712–715. http://dx.doi.org/10.1038/nclimate2742.

Westmacott, J.R., Burn, D.H., 1997. Climate change effects on the hydrologic regime within the Churchill-Nelson River Basin. J. Hydrol. 202 (1–4), 263–279. http://dx.doi.org/10.1016/S0022-1694(97)00073-5.

Westra, S., Alexander, L.V., Zwiers, F.W., 2013. Global increasing trends in annual maximum daily precipitation. J. Clim. 26 (11), 15.

Westra, S. et al., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. Rev. Geophys. 52 (3), 522–555. http://dx.doi.org/10.1002/2014RG000464.

Whitfield, P.H., 2001. Linked hydrologic and climate variations in British Columbia and Yukon. Environ. Monit. Assess. 67 (1), 217–238. http://dx.doi.org/10.1023/a:1006438723879.

Whitfield, P.H. et al., 2012. Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. Hydrol. Sci. J. 57 (8), 1562–1579. http://dx.doi.org/10.1080/02626667.2012.728706.

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 100. Academic press.

Wood, E.F. et al., 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resour. Res., 47 (5), n/a-n/a. doi: 10.1029/2010WR010090.

Yue, S., Pilon, P., Phinney, B.O.B., 2003. Canadian streamflow trend detection: impacts of serial and cross-correlation. Hydrol. Sci. J. 48 (1), 51–63. http://dx.doi.org/10.1623/hysj.48.1.51.43478.

Yue, S., Wang, C., 2002. The influence of serial correlation on the Mann-Whitney test for detecting a shift in median. Adv. Water Resour. 25 (3), 325–333. http://dx.doi.org/10.1016/S0309-1708(01)00049-5.

Zhang, X., Harvey, K.D., Hogg, W.D., Yuzyk, T.R., 2001. Trends in Canadian streamflow. Water Resour. Res. 37 (4), 987–998. http://dx.doi.org/10.1029/2000WR900357.

Zvoleff, A., 2015. gfcanalysis: Tools for Working with Hansen et al. Global Forest Change Dataset, Available online at https://cran.r-project.org/web/packages/gfcanalysis.

Appendix C – List of conference items arising from this research

Do, H.X., Gudmundsson, L., Leonard, M., Westra, S., and Senerivatne, S., 2018, The Global Streamflow Indices and Metadata (GSIM) archive – A *compilation* of 30,959 streamflow records across the globe, *AGU Fall Meeting 2018*, Washington D.C., America.

Gudmundsson L., Do, H.X., Gosling, S, Leonard, M, Liu, J, Schmied, HM, Schewe, J, Seneviratne, SI, Westra, S, Wim, T, Zhang, X and Fang Zhao, 2018, Using big data to detect and attribute global hydrological change, AGU Fall Meeting 2018, Washington D.C., America.

Do, H.X., Westra, S., Leonard, M., and Gudmundsson, L., Understanding changes in floods at the global scale and opportunity from GSIM, STAHY 2018 Workshop, Adelaide, Australia.

Gudmundsson, L., Do, H.X., Leonard, M., Westra, S., and Senerivatne, S.I, 2018, Investigating hydrological change using the Global Streamflow Indices and Metadata (GSIM) archive, EGU General Assembly 2018, Vienna, Austria.

Do, H.X., Gudmundsson, L., Leonard, M., Westra, S., and Senerivatne, S., 2018, Global-scale prediction of flood timing using atmospheric reanalysis and the Global Streamflow Indices and Metadata (GSIM) archive, EGU General Assembly 2018, Vienna, Austria

Do, H.X., Gudmundsson, L., Leonard, M., Westra, S., and Senerivatne, S.I, 2017, The Global Streamflow Indices and Metadata Archive (GSIM) – A compilation of 12 international streamflow databases, The 22nd International Congress on Modelling and Simulation, Tasmania, Australia

Do, H.X., Gudmundsson, L., Leonard, M., Westra, S., and Senerivatne, S.I., 2017, The Global Streamflow Indices and Metadata archive (G-SIM): A compilation of global streamflow time series indices and meta-data, EGU General Assembly 2017, Vienna, Austria

Do, H.X., Yu,S., 2017, Simulating Daily Streamflow with the Help of Geofabric and the AWRA-L Model, Water Information Science Symposium, Canberra, Australia.

Do, H.X., Westra, S., and Leonard, M., 2016, A global-scale investigation of trends in annual maximum streamflow, OzEWEX 3rd National Workshop, Canberra, Australia.

Do, H.X., Westra, S., and Leonard, M., 2016, Are worldwide floods changing? Evidence from global high-quality annual maximum streamflow records, EGU General Assembly 2016, Vienna, Austria.

References

ADDOR, N., NEWMAN, A. J., MIZUKAMI, N. & CLARK, M. P. 2017. The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrol. Earth Syst. Sci. Discuss.,* 2017**,** 1-31.

ALEXANDER, L., ZHANG, X., PETERSON, T., CAESAR, J., GLEASON, B., KLEIN TANK, A., HAYLOCK, M., COLLINS, D., TREWIN, B. & RAHIMZADEH, F. 2006. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres (1984–2012),* 111.

ALFIERI, L., BISSELINK, B., DOTTORI, F., NAUMANN, G., DE ROO, A., SALAMON, P., WYSER, K. & FEYEN, L. 2017. Global projections of river flood risk in a warmer world. *Earth's Future***,** n/a-n/a.

ALFIERI, L., BUREK, P., FEYEN, L. & FORZIERI, G. 2015. Global warming increases the frequency of river floods in Europe. *Hydrol. Earth Syst. Sci.,* 19**,** 2247-2260.

ARCHFIELD, S. A., CLARK, M., ARHEIMER, B., HAY, L. E., MCMILLAN, H., KIANG, J. E., SEIBERT, J., HAKALA, K., BOCK, A., WAGENER, T., FARMER, W. H., ANDRÉASSIAN, V., ATTINGER, S., VIGLIONE, A., KNIGHT, R., MARKSTROM, S. & OVER, T. 2015. Accelerating advances in continental domain hydrologic modeling. *Water Resources Research,* 51**,** 10078-10091.

ARCHFIELD, S. A., HIRSCH, R. M., VIGLIONE, A. & BLÖSCHL, G. 2016. Fragmented patterns of flood change across the United States. *Geophysical Research Letters***,** n/a-n/a.

ARNELL, N. & GOSLING, S. 2014. The impacts of climate change on river flood risk at the global scale. *Climatic Change***,** 1-15.

ARNELL, N. W. & LLOYD-HUGHES, B. 2014. The global-scale impacts of climate change on water resources and flooding under new climate and socio-economic scenarios. *Climatic Change,* 122**,** 127-140.

ARSENAULT, R., BAZILE, R., OUELLET DALLAIRE, C. & BRISSETTE, F. 2016. CANOPEX: A Canadian hydrometeorological watershed database. *Hydrological Processes,* 30**,** 2734-2736.

ASADIEH, B., KRAKAUER, N. Y. & FEKETE, B. M. 2016. Historical Trends in Mean and Extreme Runoff and Streamflow Based on Observations and Climate Models. *Water,* 8**,** 189.

ÁVILA, A., JUSTINO, F., WILSON, A., BROMWICH, D. & AMORIM, M. 2016. Recent precipitation trends, flash floods and landslides in southern Brazil. *Environmental Research Letters,* 11**,** 114029.

BARD, A., RENARD, B. & LANG, M. 2012. Floods in the Alpine Areas of Europe. *Changes in Flood Risk in Europe.* CRC Press.

BARR, I. D. & CLARK, C. D. 2012. An updated moraine map of Far NE Russia. *Journal of Maps,* 8**,** 431-436.

BECK, H. E., DE ROO, A. & VAN DIJK, A. I. J. M. 2015. Global maps of streamflow characteristics based on observations from several thousand catchments. *Journal of Hydrometeorology*.

BECK, H. E., VAN DIJK, A. I. J. M., DE ROO, A., DUTRA, E., FINK, G., ORTH, R. & SCHELLEKENS, J. 2017. Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.,* 21**,** 2881-2903.

BENNETT, B., LEONARD, M., DENG, Y. & WESTRA, S. 2018. An empirical investigation into the effect of antecedent precipitation on flood volume. *Journal of Hydrology*.

BERGHUIJS, W. R., WOODS, R. A. & HRACHOWITZ, M. 2014. A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Clim. Change,* 4**,** 583-586.

BERGHUIJS, W. R., WOODS, R. A., HUTTON, C. J. & SIVAPALAN, M. 2016. Dominant flood generating mechanisms across the United States. *Geophysical Research Letters,* 43**,** 4382-4390.

BEURTON, S. & THIEKEN, A. H. 2009. Seasonality of floods in Germany. *Hydrological Sciences Journal,* 54**,** 62-76.

BHADURI, B., BRIGHT, E., COLEMAN, P. & DOBSON, J. 2002. LandScan. *Geoinformatics,* 5**,** 34-37.

BIEMANS, H., HADDELAND, I., KABAT, P., LUDWIG, F., HUTJES, R. W. A., HEINKE, J., VON BLOH, W. & GERTEN, D. 2011. Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research,* 47.

BIERKENS, M. F. P. 2015. Global hydrology 2015: State, trends, and directions. *Water Resources Research,* 51**,** 4923-4947.

BIERKENS, M. F. P., BELL, V. A., BUREK, P., CHANEY, N., CONDON, L. E., DAVID, C. H., DE ROO, A., DÖLL, P., DROST, N., FAMIGLIETTI, J. S., FLÖRKE, M., GOCHIS, D. J., HOUSER, P., HUT, R., KEUNE, J., KOLLET, S., MAXWELL, R. M., REAGER, J. T., SAMANIEGO, L., SUDICKY, E., SUTANUDJAJA, E. H., VAN DE GIESEN, N., WINSEMIUS, H. & WOOD, E. F. 2015. Hyper-resolution global hydrological modelling: what is next? *Hydrological Processes,* 29**,** 310-320.

BLÖSCHL, G., HALL, J., PARAJKA, J., PERDIGÃO, R. A. P., MERZ, B., ARHEIMER, B., ARONICA, G. T., BILIBASHI, A., BONACCI, O., BORGA, M., ČANJEVAC, I., CASTELLARIN, A., CHIRICO, G. B., CLAPS, P., FIALA, K., FROLOVA, N., GORBACHOVA, L., GÜL, A., HANNAFORD, J., HARRIGAN, S., KIREEVA, M., KISS, A., KJELDSEN, T. R., KOHNOVÁ, S., KOSKELA, J. J., LEDVINKA, O., MACDONALD, N., MAVROVA-GUIRGUINOVA, M., MEDIERO,

L., MERZ, R., MOLNAR, P., MONTANARI, A., MURPHY, C., OSUCH, M., OVCHARUK, V., RADEVSKI, I., ROGGER, M., SALINAS, J. L., SAUQUET, E., ŠRAJ, M., SZOLGAY, J., VIGLIONE, A., VOLPI, E., WILSON, D., ZAIMI, K. & ŽIVKOVIĆ, N. 2017. Changing climate shifts timing of European floods. *Science,* 357**,** 588.

BORMANN, H., PINTER, N. & ELFERT, S. 2011. Hydrological signatures of flood trends on German rivers: Flood frequencies, flood heights and specific stages. *Journal of Hydrology,* 404**,** 50-66.

BOUWER, L. M. 2013. Projections of Future Extreme Weather Losses Under Changes in Climate and Exposure. *Risk Analysis,* 33**,** 915-930.

BRADLEY, A. A. & SMITH, J. A. 1994. The hydrometeorological environment of extreme rainstorms in the southern plains of the United States. *Journal of Applied Meteorology,* 33**,** 1418-1431.

BRADSHAW, C. J. A., SODHI, N. S., PEH, K. S. H. & BROOK, B. W. 2007. Global evidence that deforestation amplifies flood risk and severity in the developing world. *Global Change Biology,* 13**,** 2379-2395.

BRUIJNZEEL, L. A. 2004. Hydrological functions of tropical forests: not seeing the soil for the trees? *Agriculture, ecosystems & environment,* 104**,** 185-228.

BURN, D. H., HANNAFORD, J., HODGKINS, G. A., WHITFIELD, P. H., THORNE, R. & MARSH, T. 2012. Reference hydrologic networks II. Using reference hydrologic networks to assess climate-driven changes in streamflow. *Hydrological Sciences Journal,* 57**,** 1580-1593.

BURN, D. H. & WHITFIELD, P. H. 2016. Changes in floods and flood regimes in Canada. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques,* 41**,** 139-150.

BURN, D. H. & WHITFIELD, P. H. 2018. Changes in flood events inferred from centennial length streamflow data records. *Advances in Water Resources,* 121**,** 333-349.

BUSCHMAN, F. A., HOITINK, A. J. F., VAN DER VEGT, M. & HOEKSTRA, P. 2009. Subtidal water level variation controlled by river flow and tides. *Water Resources Research,* 45**,** n/a-n/a.

CANNON, A. J. 2012. Köppen versus the computer: comparing Köppen-Geiger and multivariate regression tree climate classifications in terms of climate homogeneity. *Hydrol. Earth Syst. Sci.,* 16**,** 217-229.

CHANG, H. & FRANCZYK, J. 2008. Climate Change, Land-Use Change, and Floods: Toward an Integrated Assessment. *Geography Compass,* 2**,** 1549-1579.

CHRISTEN, P. 2012. *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection*, Springer Science & Business Media.

CIESIN 2016. Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

CLARK, C. 1987. Deforestation and floods. *Environmental Conservation,* 14**,** 67-69.

CLARK, M. P., FAN, Y., LAWRENCE, D. M., ADAM, J. C., BOLSTER, D., GOCHIS, D. J., HOOPER, R. P., KUMAR, M., LEUNG, L. R., MACKAY, D. S., MAXWELL, R. M., SHEN, C., SWENSON, S. C. & ZENG, X. 2015. Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research,* 51**,** 5929-5956.

CRED 2015. The human cost of natural disasters: A global perspective. Brussels: Centre for Research on the Epidemiology of Disasters.

CUNDERLIK, J. M. & BURN, D. H. 2004. Linkages between regional trends in monthly maximum flows and selected climatic variables. *Journal of Hydrologic Engineering,* 9**,** 246-256.

CUNDERLIK, J. M., OUARDA, T. B. & BOBÉE, B. 2004. Determination of flood seasonality from hydrological records/Détermination de la saisonnalité des crues à partir de séries hydrologiques. *Hydrological Sciences Journal,* 49.

CUNDERLIK, J. M. & OUARDA, T. B. M. J. 2009. Trends in the timing and magnitude of floods in Canada. *Journal of Hydrology,* 375**,** 471-480.

DANKERS, R., ARNELL, N. W., CLARK, D. B., FALLOON, P. D., FEKETE, B. M., GOSLING, S. N., HEINKE, J., KIM, H., MASAKI, Y. & SATOH, Y. 2014. First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences,* 111**,** 3257-3261.

DEE, D. P., UPPALA, S. M., SIMMONS, A. J., BERRISFORD, P., POLI, P., KOBAYASHI, S., ANDRAE, U., BALMASEDA, M. A., BALSAMO, G., BAUER, P., BECHTOLD, P., BELJAARS, A. C. M., VAN DE BERG, L., BIDLOT, J., BORMANN, N., DELSOL, C., DRAGANI, R., FUENTES, M., GEER, A. J., HAIMBERGER, L., HEALY, S. B., HERSBACH, H., HÓLM, E. V., ISAKSEN, L., KÅLLBERG, P., KÖHLER, M., MATRICARDI, M., MCNALLY, A. P., MONGE-SANZ, B. M., MORCRETTE, J.-J., PARK, B.-K., PEUBEY, C., DE ROSNAY, P., TAVOLATO, C., THÉPAUT, J.-N. & VITART, F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society,* 137**,** 553-597.

DEGAETANO, A. T. & CASTELLANO, C. M. 2017. Future projections of extreme precipitation intensity-duration-frequency curves for climate adaptation planning in New York State. *Climate Services,* 5**,** 23-35.

DELGADO, J., APEL, H. & MERZ, B. 2010. Flood trends and variability in the Mekong river. *Hydrology and Earth System Sciences,* 14**,** 407-418.

DETTINGER, M. D. & DIAZ, H. F. 2000. Global Characteristics of Stream Flow Seasonality and Variability. *Journal of Hydrometeorology,* 1**,** 289-310.

DHAKAL, N., JAIN, S., GRAY, A., DANDY, M. & STANCIOFF, E. 2015. Nonstationarity in seasonality of extreme precipitation: A nonparametric circular statistical approach and its application. *Water Resources Research,* 51**,** 4499-4515.

DI BALDASSARRE, G. & CLAPS, P. 2011. A hydraulic study on the applicability of flood rating curves. *Hydrology Research,* 42**,** 10-19.

DO, H. X., GUDMUNDSSON, L., LEONARD, M. & WESTRA, S. 2018a. The Global Streamflow Indices and Metadata Archive - Part 1: Station catalog and Catchment boundary. PANGAEA.

DO, H. X., GUDMUNDSSON, L., LEONARD, M. & WESTRA, S. 2018b. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. *Earth Syst. Sci. Data,* 10**,** 765-785.

DO, H. X., WESTRA, S. & MICHAEL, L. 2017. A global-scale investigation of trends in annual maximum streamflow. *Journal of Hydrology*.

DONAT, M. G., ALEXANDER, L. V., YANG, H., DURRE, I., VOSE, R., DUNN, R. J. H., WILLETT, K. M., AGUILAR, E., BRUNET, M., CAESAR, J., HEWITSON, B., JACK, C., KLEIN TANK, A. M. G., KRUGER, A. C., MARENGO, J., PETERSON, T. C., RENOM, M., ORIA ROJAS, C., RUSTICUCCI, M., SALINGER, J., ELRAYAH, A. S., SEKELE, S. S., SRIVASTAVA, A. K., TREWIN, B., VILLARROEL, C., VINCENT, L. A., ZHAI, P., ZHANG, X. & KITCHING, S. 2013. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *Journal of Geophysical Research: Atmospheres,* 118**,** 2098-2118.

DOUGLAS, E., VOGEL, R. & KROLL, C. 2000. Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of Hydrology,* 240**,** 90-105.

ELVIDGE, C. D., BAUGH, K. E., ANDERSON, S. J., SUTTON, P. C. & GHOSH, T. 2012. The Night Light Development Index (NLDI): a spatially explicit measure of human development from satellite data.

FALCONE, J. A. 2011. GAGES-II: Geospatial attributes of gages for evaluating streamflow. US Geological Survey.

FEKETE, B., VÖRÖSMARTY, C. & GRABS, W. 2002a. Global Composite Runoff Fields on Observed River Discharge and Simulated Water Balances/Water System Analysis Group. University of New Hampshire, and Global Runoff Data Centre. Koblenz, Federal Institute of

Hydrology (BfG). *Koblenz, Germany, Federal Institute of Hydrology (BfG)*.

FEKETE, B. M., ROBARTS, R. D., KUMAGAI, M., NACHTNEBEL, H.-P., ODADA, E. & ZHULIDOV, A. V. 2015. Time for in situ renaissance. *Science,* 349**,** 685-686.

FEKETE, B. M., VÖRÖSMARTY, C. J. & GRABS, W. 2002b. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles,* 16**,** 15-1-15-10.

FITZHUGH, T. W. & VOGEL, R. M. 2011. The impact of dams on flood flows in the United States. *River Research and Applications,* 27**,** 1192-1215.

FORZIERI, G., FEYEN, L., RUSSO, S., VOUSDOUKAS, M., ALFIERI, L., OUTTEN, S., MIGLIAVACCA, M., BIANCHI, A., ROJAS, R. & CID, A. 2016. Multi-hazard assessment in Europe under climate change. *Climatic Change,* 137**,** 105-119.

FRANKS, S. W. & KUCZERA, G. 2002. Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales. *Water Resources Research,* 38**,** 20-1-20-7.

FREDIN, O., RUBENSDOTTER, L., VAN WELDEN, A., LARSEN, E. & LYSÅ, A. 2012. Distribution of ice marginal moraines in NW Russia. *Journal of Maps,* 8**,** 236-241.

FUCHS, T., SCHNEIDER, U. & RUDOLF, B. 2007. Global precipitation analysis products of the GPCC. *Global Precipitation Climatology Centre (GPCC). Deutscher Wetterdienst, Offenbach*.

GHASEMIZADE, M. & SCHIRMER, M. 2013. Subsurface flow contribution in the hydrological cycle: lessons learned and challenges ahead—a review. *Environmental Earth Sciences,* 69**,** 707-718.

GIUNTOLI, I., VILLARINI, G., PRUDHOMME, C. & HANNAH, D. M. 2018. Uncertainties in projected runoff over the conterminous United States. *Climatic change,* 150**,** 149-162.

GRDC 2015. Report of the Twelfth Meeting of the GRDC Steering Committee, Koblenz, Germany, 18 - 19 June 2014. *GRDC Report Series.* Koblenz, Germany: Global Runoff Data Centre (GRDC).

GREVE, P., GUDMUNDSSON, L. & SENEVIRATNE, S. I. 2018. Regional scaling of annual mean precipitation and water availability with global temperature change. *Earth Syst. Dynam.,* 9**,** 227-240.

GUDMUNDSSON, L., DO, H. X., LEONARD, M. & WESTRA, S. 2018a. The Global Streamflow Indices and Metadata Archive (GSIM) - Part 2: Time Series Indices and Homogeneity Assessment. PANGAEA.

GUDMUNDSSON, L., DO, H. X., LEONARD, M. & WESTRA, S. 2018b. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2:

Quality control, time-series indices and homogeneity assessment. *Earth Syst. Sci. Data,* 10**,** 787-804.

GUDMUNDSSON, L., DO, H. X., LEONARD, M., WESTRA, S. & SENEVIRATNE, S. I. in review. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality Control, Time-series Indices and Homogeneity Assessment. *Earth System Science Data Discussions*.

GUDMUNDSSON, L., LEONARD, M., DO, H. X., WESTRA, S. & SENEVIRATNE, S. I. 2019. Observed Trends in Global Indicators of Mean and Extreme Streamflow. *Geophysical Research Letters,* 46.

GUDMUNDSSON, L. & SENEVIRATNE, S. I. 2015. Towards observation-based gridded runoff estimates for Europe. *Hydrology and Earth System Sciences,* 19**,** 2859-2879.

GUDMUNDSSON, L. & SENEVIRATNE, S. I. 2016. Observational gridded runoff estimates for Europe (E-RUN version 1.0). *Earth Syst. Sci. Data Discuss.,* 2016**,** 1-27.

GUDMUNDSSON, L., SENEVIRATNE, S. I. & ZHANG, X. 2017. Anthropogenic climate change detected in European renewable freshwater resources. *Nature Climate Change,* 7**,** 813.

GUDMUNDSSON, L., TALLAKSEN, L. M., STAHL, K., CLARK, D. B., DUMONT, E., HAGEMANN, S., BERTRAND, N., GERTEN, D., HEINKE, J., HANASAKI, N., VOSS, F. & KOIRALA, S. 2012a. Comparing Large-Scale Hydrological Model Simulations to Observed Runoff Percentiles in Europe. *Journal of Hydrometeorology,* 13**,** 604-620.

GUDMUNDSSON, L., WAGENER, T., TALLAKSEN, L. M. & ENGELAND, K. 2012b. Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. *Water Resources Research,* 48**,** n/a-n/a.

GUERREIRO, S. B., FOWLER, H. J., BARBERO, R., WESTRA, S., LENDERINK, G., BLENKINSOP, S., LEWIS, E. & LI, X.-F. 2018. Detection of continental-scale intensification of hourly rainfall extremes. *Nature Climate Change,* 8**,** 803-807.

GUHA-SAPIR, D., HOYOIS, P. & BELOW, R. 2015. Annual Disaster Statistical Review 2014: The numbers and trends. UCL.

GUHA-SAPIR, D., HOYOIS, P., AND BELOW, R. 2014. Annual Disaster Statistical Review 2013: The Numbers and Trends. Brussels: CRED.

GUIMBERTEAU, M., DUCHARNE, A., CIAIS, P., BOISIER, J.-P., PENG, S., DE WEIRDT, M. & VERBEECK, H. 2014. Testing conceptual and physically based soil hydrology schemes against observations for the Amazon Basin. *Geoscientific Model Development,* 7**,** 1115-1136.

GUPTA, H., PERRIN, C., BLOSCHL, G., MONTANARI, A., KUMAR, R., CLARK, M. & ANDRÉASSIAN, V. 2014. Large-sample hydrology: a

need to balance depth with breadth. *Hydrology and Earth System Sciences,* 18**,** p. 463-p. 477.

HALL, J., ARHEIMER, B., BORGA, M., BRÁZDIL, R., CLAPS, P., KISS, A., KJELDSEN, T. R., KRIAUČIŪNIENĖ, J., KUNDZEWICZ, Z. W., LANG, M., LLASAT, M. C., MACDONALD, N., MCINTYRE, N., MEDIERO, L., MERZ, B., MERZ, R., MOLNAR, P., MONTANARI, A., NEUHOLD, C., PARAJKA, J., PERDIGÃO, R. A. P., PLAVCOVÁ, L., ROGGER, M., SALINAS, J. L., SAUQUET, E., SCHÄR, C., SZOLGAY, J., VIGLIONE, A. & BLÖSCHL, G. 2014. Understanding flood regime changes in Europe: a state-of-the-art assessment. *Hydrol. Earth Syst. Sci.,* 18**,** 2735-2772.

HALL, J. & BLÖSCHL, G. 2018. Spatial patterns and characteristics of flood seasonality in Europe. *Hydrol. Earth Syst. Sci.,* 22**,** 3883-3901.

HALLEGATTE, S., GREEN, C., NICHOLLS, R. J. & CORFEE-MORLOT, J. 2013. Future flood losses in major coastal cities. *Nature Clim. Change,* 3**,** 802-806.

HANASAKI, N., KANAE, S., OKI, T., MASUDA, K., MOTOYA, K., SHIRAKAWA, N., SHEN, Y. & TANAKA, K. 2008a. An integrated model for the assessment of global water resources–Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences,* 12**,** 1007-1025.

HANASAKI, N., KANAE, S., OKI, T., MASUDA, K., MOTOYA, K., SHIRAKAWA, N., SHEN, Y. & TANAKA, K. 2008b. An integrated model for the assessment of global water resources–Part 2: Applications and assessments. *Hydrology and Earth System Sciences,* 12**,** 1027-1037.

HANNAFORD, J., BUYS, G., STAHL, K. & TALLAKSEN, L. M. 2013. The influence of decadal-scale variability on trends in long European streamflow records. *Hydrol. Earth Syst. Sci.,* 17**,** 2717-2733.

HANNAFORD, J. & MARSH, T. 2006. An assessment of trends in UK runoff and low flows using a network of undisturbed catchments. *International Journal of Climatology,* 26**,** 1237-1253.

HANNAFORD, J. & MARSH, T. J. 2008. High-flow and flood trends in a network of undisturbed catchments in the UK. *International Journal of Climatology,* 28**,** 1325-1338.

HANNAH, D. M., DEMUTH, S., VAN LANEN, H. A. J., LOOSER, U., PRUDHOMME, C., REES, G., STAHL, K. & TALLAKSEN, L. M. 2011. Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes,* 25**,** 1191-1200.

HANSEN, M. C., POTAPOV, P. V., MOORE, R., HANCHER, M., TURUBANOVA, S. A., TYUKAVINA, A., THAU, D., STEHMAN, S. V., GOETZ, S. J., LOVELAND, T. R., KOMMAREDDY, A., EGOROV, A., CHINI, L., JUSTICE, C. O. & TOWNSHEND, J. R. G.

2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science,* 342**,** 850-853.

HANSON, S., NICHOLLS, R., RANGER, N., HALLEGATTE, S., CORFEE-MORLOT, J., HERWEIJER, C. & CHATEAU, J. 2011. A global ranking of port cities with high exposure to climate extremes. *Climatic change,* 104**,** 89-111.

HARTMANN, J. & MOOSDORF, N. 2012. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems,* 13.

HE, J. & SODEN, B. J. 2016. The impact of SST biases on projections of anthropogenic climate change: A greater role for atmosphere-only models? *Geophysical Research Letters,* 43**,** 7745-7750.

HENCK, A. C., MONTGOMERY, D. R., HUNTINGTON, K. W. & LIANG, C. 2010. Monsoon control of effective discharge, Yunnan and Tibet. *Geology,* 38**,** 975-978.

HENGL, T., MENDES DE JESUS, J., HEUVELINK, G. B. M., RUIPEREZ GONZALEZ, M., KILIBARDA, M., BLAGOTIĆ, A., SHANGGUAN, W., WRIGHT, M. N., GENG, X., BAUER-MARSCHALLINGER, B., GUEVARA, M. A., VARGAS, R., MACMILLAN, R. A., BATJES, N. H., LEENAARS, J. G. B., RIBEIRO, E., WHEELER, I., MANTEL, S. & KEMPEN, B. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE,* 12**,** e0169748.

HERSCHY, R. 1994. The analysis of uncertainties in the stage-discharge relation. *Flow Measurement and Instrumentation,* 5**,** 188-190.

HEWLETT, J. D. & HIBBERT, A. R. 1967. Factors affecting the response of small watersheds to precipitation in humid areas.

HIRABAYASHI, Y., MAHENDRAN, R., KOIRALA, S., KONOSHIMA, L., YAMAZAKI, D., WATANABE, S., KIM, H. & KANAE, S. 2013. Global flood risk under climate change. *Nature Clim. Change,* 3**,** 816-821.

HOCK, R. 2003. Temperature index melt modelling in mountain areas. *Journal of Hydrology,* 282**,** 104-115.

HODGKINS, G. A., WHITFIELD, P. H., BURN, D. H., HANNAFORD, J., RENARD, B., STAHL, K., FLEIG, A. K., MADSEN, H., MEDIERO, L., KORHONEN, J., MURPHY, C. & WILSON, D. 2017. Climate-driven variability in the occurrence of major floods across North America and Europe. *Journal of Hydrology,* 552**,** 704-717.

HOEGH-GULDBERG, O., JACOB, D., TAYLOR, M., BINDI, M., BROWN, S., CAMILLONI, I., DIEDHIOU, A., DJALANTE, R., EBI, K. & ENGELBRECHT, F. 2018. Impacts of 1.5 ºC global warming on natural and human systems.

HUANG, S., KUMAR, R., FLÖRKE, M., YANG, T., HUNDECHA, Y., KRAFT, P., GAO, C., GELFAN, A., LIERSCH, S., LOBANOVA, A.,

STRAUCH, M., VAN OGTROP, F., REINHARDT, J., HABERLANDT, U. & KRYSANOVA, V. 2016. Evaluation of an ensemble of regional hydrological models in 12 large-scale river basins worldwide. *Climatic Change***,** 1-17.

HUNGER, M. & DÖLL, P. 2008. Value of river discharge data for global-scale hydrological modeling. *Hydrology and Earth System Sciences Discussions,* 12**,** 841-861.

ILIOPOULOU, T., AGUILAR, C., ARHEIMER, B., BERMÚDEZ, M., BEZAK, N., FICCHÌ, A., KOUTSOYIANNIS, D., PARAJKA, J., POLO, M. J., THIREL, G. & MONTANARI, A. 2019. A large sample analysis of European rivers on seasonal river flow correlation and its physical drivers. *Hydrol. Earth Syst. Sci.,* 23**,** 73-91.

INGLE SMITH, D. 1999. Floods: physical processes and human impacts by K. Smith and R. Ward, John Wiley, Chichester 1998. No. of pages: 382. *Earth Surface Processes and Landforms,* 24**,** 1261-1261.

IPCC 2012. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation,* Cambridge, UK, and New York, NY, USA, Cambridge University Press.

IPCC 2014. Summary for policymakers. *In:* FIELD, C. B., V.R. BARROS, D.J. DOKKEN, K.J. MACH, M.D. MASTRANDREA, T.E. BILIR, M. CHATTERJEE, K.L. EBI, Y.O. ESTRADA, R.C. GENOVA, B. GIRMA, E.S. KISSEL, A.N. LEVY, S. MACCRACKEN, P.R. MASTRANDREA & WHITE, L. L. (eds.) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

ISHAK, E., RAHMAN, A., WESTRA, S., SHARMA, A. & KUCZERA, G. 2013. Evaluating the non-stationarity of Australian annual maximum flood. *Journal of Hydrology,* 494**,** 134-145.

IVANCIC, T. & SHAW, S. 2015. Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. *Climatic Change***,** 1-13.

JARAMILLO, F. & DESTOUNI, G. 2015. Local flow regulation and irrigation raise global human water consumption and footprint. *Science,* 350**,** 1248-1251.

JENSON, S. K. & DOMINGUE, J. O. 1988. Extracting topographic structure from digital elevation data for geographic information-system analysis. *Photogrammetric Engineering and Remote Sensing,* 54**,** 1593-1600.

JOHNSON, F., WHITE, C. J., VAN DIJK, A., EKSTROM, M., EVANS, J. P., JAKOB, D., KIEM, A. S., LEONARD, M., ROUILLARD, A. & WESTRA, S. 2016. Natural hazards in Australia: floods. *Climatic Change***,** 1-15.

KEENAN, R. J., REAMS, G. A., ACHARD, F., DE FREITAS, J. V., GRAINGER, A. & LINDQUIST, E. 2015. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management,* 352**,** 9-20.

KETTNER, A. J., COHEN, S., OVEREEM, I., FEKETE, B. M., BRAKENRIDGE, G. R. & SYVITSKI, J. P. 2018. Estimating Change in Flooding for the 21st Century Under a Conservative RCP Forcing: A Global Hydrological Modeling Assessment. *Global Flood Hazard: Applications in Modeling, Mapping, and Forecasting***,** 157-167.

KHARIN, V. V., ZWIERS, F. W., ZHANG, X. & HEGERL, G. C. 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *Journal of Climate,* 20**,** 1419-1444.

KIKTEV, D., CAESAR, J., ALEXANDER, L. V., SHIOGAMA, H. & COLLIER, M. 2007. Comparison of observed and multimodeled trends in annual extremes of temperature and precipitation. *Geophysical research letters,* 34.

KIKTEV, D., SEXTON, D. M., ALEXANDER, L. & FOLLAND, C. K. 2003. Comparison of modeled and observed trends in indices of daily climate extremes. *Journal of Climate,* 16**,** 3560-3571.

KIM, H. 2017. Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1). Data Integration and Analysis System (DIAS).

KOUSKY, C. 2014. Informing climate adaptation: A review of the economic costs of natural disasters. *Energy Economics,* 46**,** 576-592.

KOZLOWSKI, T. T. 1984. CHAPTER 1 - Extent, Causes, and Impacts of Flooding. *In:* KOZLOWSKI, T. T. (ed.) *Flooding and Plant Growth.* San Diego: Academic Press.

KRON, W. 2005. Flood Risk = Hazard • Values • Vulnerability. *Water International,* 30**,** 58-68.

KUMAR, S., MERWADE, V., KINTER III, J. L. & NIYOGI, D. 2013. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. *Journal of Climate,* 26**,** 4168-4185.

KUNDZEWICZ, Z. W., GRACZYK, D., MAURER, T., PRZYMUSIŃSKA, I., RADZIEJEWSKI, M., SVENSSON, C. & SZWED, M. 2004. Detection of change in world-wide hydrological time series of maximum annual flow. *In:* KUNDZEWICZ, Z. W. (ed.) *GRDC Report Series.* Koblenz, Germany: Global Runoff Date Centre.

KUNDZEWICZ, Z. W., KANAE, S., SENEVIRATNE, S. I., HANDMER, J., NICHOLLS, N., PEDUZZI, P., MECHLER, R., BOUWER, L. M., ARNELL, N., MACH, K., MUIR-WOOD, R., BRAKENRIDGE, G. R., KRON, W., BENITO, G., HONDA, Y., TAKAHASHI, K. &

SHERSTYUKOV, B. 2013. Flood risk and climate change: global and regional perspectives. *Hydrological Sciences Journal,* 59**,** 1-28.

KUNDZEWICZ, Z. W. & MONDIALE, O. M. 2000. *Detecting trend and other changes in hydrological data*, WMO.

KUNDZEWICZ, Z. W., PLATE, E. J., RODDA, H. J., RODDA, J. C., SCHELLNHUBER, H. J. & STRUPCZEWSKI, W. G. 2012. Changes in flood risk–setting the stage. *Changes in flood risk in Europe.* CRC Press.

LAJOIE, F., ASSANI, A. A., ROY, A. G. & MESFIOUI, M. 2007. Impacts of dams on monthly flow characteristics. The influence of watershed size and seasons. *Journal of hydrology,* 334**,** 423-439.

LEE, D., WARD, P. & BLOCK, P. 2015. Defining high-flow seasons using temporal streamflow patterns from a global model. *Hydrol. Earth Syst. Sci.,* 19**,** 4689-4705.

LEHMANN, J., COUMOU, D. & FRIELER, K. 2015. Increased record-breaking precipitation events under global warming. *Climatic Change***,** 1-15.

LEHNER, B. 2012. Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network.

LEHNER, B. & GRILL, G. 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes,* 27**,** 2171-2186.

LEHNER, B., LIERMANN, C. R., REVENGA, C., VÖRÖSMARTY, C., FEKETE, B., CROUZET, P., DÖLL, P., ENDEJAN, M., FRENKEN, K. & MAGOME, J. 2011. Global reservoir and dam (grand) database. *Technical Documentation, Version,* 1.

LEHNER, B., VERDIN, K. & JARVIS, A. 2006. HydroSHEDS technical documentation, version 1.0. *World Wildlife Fund US, Washington, DC***,** 1-27.

LEHNER, B., VERDIN, K. & JARVIS, A. 2008. New Global Hydrography Derived From Spaceborne Elevation Data. *Eos, Transactions American Geophysical Union,* 89**,** 93-94.

LEONARD, M., METCALFE, A. & LAMBERT, M. 2008. Frequency analysis of rainfall and streamflow extremes accounting for seasonal and climatic partitions. *Journal of hydrology,* 348**,** 135-147.

LINDSAY, J. B., ROTHWELL, J. J. & DAVIES, H. 2008. Mapping outlet points used for watershed delineation onto DEM-derived stream networks. *Water resources research,* 44.

LINS, H. F. & SLACK, J. R. 1999. Streamflow trends in the United States. *Geophysical research letters,* 26**,** 227-230.

LIU, X., TANG, Q., CUI, H., MU, M., GERTEN, D., GOSLING, S. N., MASAKI, Y., SATOH, Y. & WADA, Y. 2017. Multimodel uncertainty

changes in simulated river flows induced by human impact parameterizations. *Environmental Research Letters,* 12**,** 025009.

LORENZ, R., HERGER, N., SEDLÁČEK, J., EYRING, V., FISCHER, E. M. & KNUTTI, R. 2018. Prospects and caveats of weighting climate models for summer maximum temperature projections over North America. *Journal of Geophysical Research: Atmospheres,* 123**,** 4509-4526.

LUCAS, C. & TINGLEY, D. 2016. translateR. Available online at https://cran.r-project.org/web/packages/translateR/index.html.

MALLAKPOUR, I. & VILLARINI, G. 2015. The changing nature of flooding across the central United States. *Nature Clim. Change,* 5**,** 250-254.

MANGINI, W., VIGLIONE, A., HALL, J., HUNDECHA, Y., CEOLA, S., MONTANARI, A., ROGGER, M., SALINAS, J. L., BORZÌ, I. & PARAJKA, J. 2018. Detection of trends in magnitude and frequency of flood peaks across Europe. *Hydrological Sciences Journal,* 63**,** 493-512.

MARDIA, K. V. & JUPP, P. E. 2009. *Directional statistics*, John Wiley & Sons.

MARENGO, J. A. 2006. On the hydrological cycle of the Amazon Basin: A historical review and current state-of-the-art. *Revista Brasileira de Meteorologia,* 21**,** 1-19.

MARTHEWS, T., DADSON, S., LEHNER, B., ABELE, S. & GEDNEY, N. 2015. High-resolution global topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences,* 19**,** 91-104.

MEDIERO, L., KJELDSEN, T. R., MACDONALD, N., KOHNOVA, S., MERZ, B., VOROGUSHYN, S., WILSON, D., ALBURQUERQUE, T., BLÖSCHL, G., BOGDANOWICZ, E., CASTELLARIN, A., HALL, J., KOBOLD, M., KRIAUCIUNIENE, J., LANG, M., MADSEN, H., ONUŞLUEL GÜL, G., PERDIGÃO, R. A. P., ROALD, L. A., SALINAS, J. L., TOUMAZIS, A. D., VEIJALAINEN, N. & ÞÓRARINSSON, Ó. 2015. Identification of coherent flood regions across Europe by using the longest streamflow records. *Journal of Hydrology,* 528**,** 341-360.

MERZ, B., DUNG, N. V., APEL, H., GERLITZ, L., SCHRÖTER, K., STEIROU, E. & VOROGUSHYN, S. 2018. Spatial coherence of flood-rich and flood-poor periods across Germany. *Journal of Hydrology,* 559**,** 813-826.

MERZ, B., KUNDZEWICZ, Z., DELGADO, J., HUNDECHA, Y. & KREIBICH, H. 2012a. Detection and attribution of changes in flood hazard and risk. *Changes in flood risk in Europe. IAHS Special Publication,* 10**,** 435-458.

MERZ, B., VOROGUSHYN, S., UHLEMANN, S., DELGADO, J. & HUNDECHA, Y. 2012b. HESS Opinions" More efforts and scientific rigour are needed to attribute trends in flood time series". *Hydrology and Earth System Sciences,* 16**,** 1379-1387.

MIAO, Q. 2018. Are We Adapting to Floods? Evidence from Global Flooding Fatalities. *Risk Analysis,* 0.

MICEVSKI, T., FRANKS, S. W. & KUCZERA, G. 2006. Multidecadal variability in coastal eastern Australian flood data. *Journal of Hydrology,* 327**,** 219-225.

MILLS, E. 2005. Insurance in a climate of change. *Science,* 309**,** 1040+.

MILLY, P. C. D., BETANCOURT, J., FALKENMARK, M., HIRSCH, R. M., KUNDZEWICZ, Z. W., LETTENMAIER, D. P., STOUFFER, R. J., DETTINGER, M. D. & KRYSANOVA, V. 2015. On Critiques of "Stationarity is Dead: Whither Water Management?". *Water Resources Research,* 51**,** 7785-7789.

MILLY, P. C. D., WETHERALD, R. T., DUNNE, K. & DELWORTH, T. L. 2002. Increasing risk of great floods in a changing climate. *Nature,* 415**,** 514-517.

MIN, S.-K., ZHANG, X., ZWIERS, F. W. & HEGERL, G. C. 2011. Human contribution to more-intense precipitation extremes. *Nature,* 470**,** 378-381.

MITCHELL, T. D. & JONES, P. D. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology,* 25**,** 693-712.

MORÁN-TEJEDA, E., LÓPEZ-MORENO, J. I., VICENTE-SERRANO, S. M., LORENZO-LACRUZ, J. & CEBALLOS-BARBANCHO, A. 2012. The contrasted evolution of high and low flows and precipitation indices in the Duero basin (Spain). *Hydrological Sciences Journal,* 57**,** 591-611.

MUELLER SCHMIED, H., ADAM, L., EISNER, S., FINK, G., FLÖRKE, M., KIM, H., OKI, T., PORTMANN, F. T., REINECKE, R. & RIEDEL, C. 2016. Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrology and Earth System Sciences,* 20**,** 2877-2898.

MÜLLER SCHMIED, H., EISNER, S., FRANZ, D., WATTENBACH, M., PORTMANN, F. T., FLÖRKE, M. & DÖLL, P. 2014. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences,* 18**,** 3511-3538.

MUNICH RE 2004. Megacities – Megarisks Trends and challenges for insurance and risk management. Germany: Munich Re Group.

MUNICH RE 2015. NatCatSERVICE: Loss events worldwide 1980 – 2014 Munich: Munich Re.

NAJIBI, N. & DEVINENI, N. 2018. Recent trends in the frequency and duration of global floods. *Earth Syst. Dynam.,* 9**,** 757-783.

NELSON, B. 2009. Data sharing: Empty archives. *Nature,* 461.

NKA, B. N., OUDIN, L., KARAMBIRI, H., PATUREL, J. E. & RIBSTEIN, P. 2015. Trends in floods in West Africa: analysis based on 11 catchments in the region. *Hydrol. Earth Syst. Sci.,* 19**,** 4707-4719.

PADRÓN, R. S., GUDMUNDSSON, L. & SENEVIRATNE, S. I. 2019. Observational Constraints Reduce Likelihood of Extreme Changes in Multidecadal Land Water Availability. *Geophysical Research Letters,* 46.

PALL, P., AINA, T., STONE, D. A., STOTT, P. A., NOZAWA, T., HILBERTS, A. G. J., LOHMANN, D. & ALLEN, M. R. 2011. Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature,* 470**,** 382-385.

PARAJKA, J., KOHNOVÁ, S., BÁLINT, G., BARBUC, M., BORGA, M., CLAPS, P., CHEVAL, S., DUMITRESCU, A., GAUME, E., HLAVČOVÁ, K., MERZ, R., PFAUNDLER, M., STANCALIE, G., SZOLGAY, J. & BLÖSCHL, G. 2010. Seasonal characteristics of flood regimes across the Alpine–Carpathian range. *Journal of Hydrology,* 394**,** 78-89.

PATHIRAJA, S., WESTRA, S. & SHARMA, A. 2012. Why continuous simulation? The role of antecedent moisture in design flood estimation. *Water Resources Research,* 48**,** n/a-n/a.

PEDUZZI, P., DAO, H., HEROLD, C. & MOUTON, F. 2009. Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index. *Nat. Hazards Earth Syst. Sci.,* 9**,** 1149-1159.

PETROW, T. & MERZ, B. 2009. Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002. *Journal of Hydrology,* 371**,** 129-141.

PILGRIM, E., INSTITUTION OF ENGINEERS, A., PILGRIM, D. & CANTERFORD, R. 1987. *Australian rainfall and runoff*, Institution of Engineers, Australia.

POKHREL, Y., HANASAKI, N., KOIRALA, S., CHO, J., YEH, P. J.-F., KIM, H., KANAE, S. & OKI, T. 2012. Incorporating Anthropogenic Water Regulation Modules into a Land Surface Model. *Journal of Hydrometeorology,* 13**,** 255-269.

POLITIS, D. N. 2003. The impact of bootstrap methods on time series analysis. *Statistical Science,* 18**,** 219-230.

POLITIS, D. N. & WHITE, H. 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews,* 23**,** 53-70.

RAO, V. B. & HADA, K. 1990. Characteristics of rainfall over Brazil: Annual variations and connections with the Southern Oscillation. *Theoretical and Applied Climatology,* 42**,** 81-91.

ROST, S., GERTEN, D., BONDEAU, A., LUCHT, W., ROHWER, J. & SCHAPHOFF, S. 2008. Agricultural green and blue water consumption and its influence on the global water system. *Water Resources Research,* 44.

RUBEL, F. & KOTTEK, M. 2010. Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorologische Zeitschrift,* 19**,** 135-141.

SAKAGUCHI, K., ZENG, X. & BRUNKE, M. A. 2012. Temporal-and spatial-scale dependence of three CMIP3 climate models in simulating the surface temperature trend in the twentieth century. *Journal of Climate,* 25**,** 2456-2470.

SCHAAKE, J., CONG, S. & DUAN, Q. 2006. The US MOPEX data set. *IAHS publication,* 307**,** 9.

SCHILLING, K. E. & LIBRA, R. D. 2003. INCREASED BASEFLOW IN IOWA OVER THE SECOND HALF OF THE 20TH CENTURY1. *JAWRA Journal of the American Water Resources Association,* 39**,** 851-860.

SCHNEIDER, A., JOST, A., COULON, C., SILVESTRE, M., THÉRY, S. & DUCHARNE, A. 2017. Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density. *Geophysical Research Letters***,** n/a-n/a.

SENEVIRATNE, S. I., NICHOLLS, N., EASTERLING, D., GOODESS, C. M., KANAE, S., KOSSIN, J., LUO, Y., MARENGO, J., MCINNES, K. & RAHIMI, M. 2012. Changes in climate extremes and their impacts on the natural physical environment. *Managing the risks of extreme events and disasters to advance climate change adaptation***,** 109-230.

SHARMA, A., WASKO, C. & LETTENMAIER, D. P. 2018. If Precipitation Extremes Are Increasing, Why Aren't Floods? *Water Resources Research,* 0.

SHENG, Y. & LINGHUI, W. 2012. Detection of Changes. *Changes in Flood Risk in Europe.* CRC Press.

SIEBERT, S., KUMMU, M., PORKKA, M., DÖLL, P., RAMANKUTTY, N. & SCANLON, B. R. 2015. A global data set of the extent of irrigated land from 1900 to 2005. *Hydrol. Earth Syst. Sci.,* 19**,** 1521-1545.

SIL, A. & SITHARAM, T. 2016. Detection of Local Site Conditions in Tripura and Mizoram Using the Topographic Gradient Extracted from Remote Sensing Data and GIS Techniques. *Natural Hazards Review,* 18**,** 04016009.

SLATER, L. J., SINGER, M. B. & KIRCHNER, J. W. 2015. Hydrologic versus geomorphic drivers of trends in flood hazard. *Geophysical Research Letters,* 42**,** 370-376.

SLATER, L. J. & VILLARINI, G. 2016. Recent trends in U.S. flood risk. *Geophysical Research Letters,* 43**,** 12,428-12,436.

SMITH, J. A., VILLARINI, G. & BAECK, M. L. 2011. Mixture Distributions and the Hydroclimatology of Extreme Rainfall and Flooding in the Eastern United States. *Journal of Hydrometeorology,* 12**,** 294-309.

SMITH, K. 2003. *Environmental hazards: assessing risk and reducing disaster*, Routledge.

SOOD, A. & SMAKHTIN, V. 2015. Global hydrological models: a review. *Hydrological Sciences Journal,* 60**,** 549-565.

STACKE, T. & HAGEMANN, S. 2012. Development and evaluation of a global dynamical wetlands extent scheme. *Hydrology and Earth System Sciences,* 16**,** 2915.

STAHL, K., HISDAL, H., HANNAFORD, J., TALLAKSEN, L., VAN LANEN, H., SAUQUET, E., DEMUTH, S., FENDEKOVA, M. & JORDAR, J. 2010. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. *Hydrology and Earth System Sciences,* 14**,** p. 2367-p. 2382.

STAHL, K., TALLAKSEN, L., HANNAFORD, J. & VAN LANEN, H. 2011. „Abflusstrends in Europa: Vergleich eines Multi-Modell-Experiments mit Beobachtungen ". *Blöschl, G. und Merz, R.(Hrsg.): Hydrologie & Wasserwirtschaft-von der Theorie zur Praxis. Beiträge zum Tag der Hydrologie,* 24**,** 25.

STEVENS, A. J., CLARKE, D. & NICHOLLS, R. J. 2016. Trends in reported flooding in the UK: 1884–2013. *Hydrological Sciences Journal,* 61**,** 50-63.

STEVENSON, S. N. & SCHUMACHER, R. S. 2014. A 10-year survey of extreme rainfall events in the central and eastern United States using gridded multisensor precipitation analyses. *Monthly Weather Review,* 142**,** 3147-3162.

STOVER, S. C. & MONTGOMERY, D. R. 2001. Channel change and flooding, Skokomish River, Washington. *Journal of Hydrology,* 243**,** 272-286.

SUN, Q., MIAO, C., DUAN, Q., ASHOURI, H., SOROOSHIAN, S. & HSU, K.-L. 2018. A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics,* 56**,** 79-107.

SWISS RE 2015. Natural catastropes and man-made disaster in 2014. *Sigma.* Zurich, Switzerland: Swiss Reinsurance Company.

TEIXEIRA, M. D. S. & SATYAMURTY, P. 2011. Trends in the Frequency of Intense Precipitation Events in Southern and Southeastern Brazil during 1960–2004. *Journal of Climate,* 24**,** 1913-1921.

TURNER, M., BARI, M., AMIRTHANATHAN, G. & AHMAD, Z. 2012. Australian network of hydrologic reference stations-advances in design, development and implementation.

VAN BEEK, L., WADA, Y. & BIERKENS, M. F. 2011. Global monthly water stress: 1. Water balance and water availability. *Water Resources Research,* 47.

VAN DIJK, A. I. J. M., PEÑA-ARANCIBIA, J. L., WOOD, E. F., SHEFFIELD, J. & BECK, H. E. 2013. Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide. *Water Resources Research,* 49**,** 2729-2746.

VAN DIJK, A. I. J. M., VAN NOORDWIJK, M., CALDER, I. R., BRUIJNZEEL, S. L. A., SCHELLEKENS, J. & CHAPPELL, N. A. 2009. Forest–flood relation still tenuous – comment on 'Global evidence that deforestation amplifies flood risk and severity in the developing world' by C. J. A. Bradshaw, N.S. Sodi, K. S.-H. Peh and B.W. Brook. *Global Change Biology,* 15**,** 110-115.

VELDKAMP, T. I. E., ZHAO, F., WARD, P. J., DE MOEL, H., AERTS, J. C., SCHMIED, H. M., PORTMANN, F. T., MASAKI, Y., POKHREL, Y. & LIU, X. 2018. Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study. *Environmental Research Letters,* 13**,** 055008.

VILLARINI, G. 2016. On the seasonality of flooding across the continental United States. *Advances in Water Resources,* 87**,** 80-91.

VILLARINI, G., GOSKA, R., SMITH, J. A. & VECCHI, G. A. 2014. North Atlantic tropical cyclones and US flooding. *Bulletin of the American Meteorological Society,* 95**,** 1381-1388.

VILLARINI, G., SERINALDI, F., SMITH, J. A. & KRAJEWSKI, W. F. 2009. On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resources Research,* 45**,** n/a-n/a.

VILLARINI, G. & SMITH, J. A. 2010. Flood peak distributions for the eastern United States. *Water Resources Research,* 46.

VON BLOH, W., ROST, S., GERTEN, D. & LUCHT, W. 2010. Efficient parallelization of a dynamic global vegetation model with river routing. *Environmental Modelling & Software,* 25**,** 685-690.

VON STORCH, H. & NAVARRA, A. 2013. *Analysis of climate variability: Applications of statistical techniques proceedings of an autumn school organized by the Commission of the European Community on Elba from October 30 to November 6, 1993*, Springer Science & Business Media.

VÖRÖSMARTY, C. J., MOORE, B., GRACE, A. L., GILDEA, M. P., MELILLO, J. M., PETERSON, B. J., RASTETTER, E. B. & STEUDLER, P. A. 1989. Continental scale models of water balance and fluvial transport: an application to South America. *Global biogeochemical cycles,* 3**,** 241-265.

WADA, Y., VAN BEEK, L. & BIERKENS, M. F. 2011. Modelling global water stress of the recent past: on the relative importance of trends in water demand and climate variability. *Hydrology and Earth System Sciences,* 15**,** 3785-3805.

WANDERS, N. & WADA, Y. 2015. Decadal predictability of river discharge with climate oscillations over the 20th and early 21st century. *Geophysical Research Letters,* 42**,** 10,689-10,695.

WARD, P. J., EISNER, S., FLÖRKE, M., DETTINGER, M. D. & KUMMU, M. 2014. Annual flood sensitivities to El Niño–Southern Oscillation at the global scale. *Hydrol. Earth Syst. Sci.,* 18**,** 47-66.

WARD, P. J., JONGMAN, B., SALAMON, P., SIMPSON, A., BATES, P., DE GROEVE, T., MUIS, S., DE PEREZ, E. C., RUDARI, R., TRIGG, M. A. & WINSEMIUS, H. C. 2015. Usefulness and limitations of global flood risk models. *Nature Clim. Change,* 5**,** 712-715.

WARD, P. J., JONGMAN, B., WEILAND, F. S., BOUWMAN, A., VAN BEEK, R., BIERKENS, M. F., LIGTVOET, W. & WINSEMIUS, H. C. 2013. Assessing flood risk at the global scale: model setup, results, and sensitivity. *Environmental research letters,* 8**,** 044019.

WASKO, C. & SHARMA, A. 2017. Global assessment of flood and storm extremes with increased temperatures. *Scientific Reports,* 7**,** 7945.

WESTMACOTT, J. R. & BURN, D. H. 1997. Climate change effects on the hydrologic regime within the Churchill-Nelson River Basin. *Journal of Hydrology,* 202**,** 263-279.

WESTRA, S., ALEXANDER, L. V. & ZWIERS, F. W. 2013. Global Increasing Trends in Annual Maximum Daily Precipitation. *Journal of Climate,* 26**,** 15.

WESTRA, S., FOWLER, H. J., EVANS, J. P., ALEXANDER, L. V., BERG, P., JOHNSON, F., KENDON, E. J., LENDERINK, G. & ROBERTS, N. M. 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics,* 52**,** 522-555.

WHITFIELD, P. H. 2001. Linked Hydrologic and Climate Variations in British Columbia and Yukon. *Environmental Monitoring and Assessment,* 67**,** 217-238.

WHITFIELD, P. H., BURN, D. H., HANNAFORD, J., HIGGINS, H., HODGKINS, G. A., MARSH, T. & LOOSER, U. 2012. Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal,* 57**,** 1562-1579.

WILKS, D. S. 2011. *Statistical methods in the atmospheric sciences*, Academic press.

WINSEMIUS, H. C., VAN BEEK, L. P. H., JONGMAN, B., WARD, P. J. & BOUWMAN, A. 2013. A framework for global river flood risk assessments. *Hydrol. Earth Syst. Sci.,* 17**,** 1871-1892.

WOLDEMESKEL, F. & SHARMA, A. 2016. Should flood regimes change in a warming climate? The role of antecedent moisture conditions. *Geophysical Research Letters,* 43**,** 7556-7563.

WOOD, E. F., ROUNDY, J. K., TROY, T. J., VAN BEEK, L. P. H., BIERKENS, M. F. P., BLYTH, E., DE ROO, A., DÖLL, P., EK, M., FAMIGLIETTI, J., GOCHIS, D., VAN DE GIESEN, N., HOUSER, P., JAFFÉ, P. R., KOLLET, S., LEHNER, B., LETTENMAIER, D. P., PETERS-LIDARD, C., SIVAPALAN, M., SHEFFIELD, J., WADE, A. & WHITEHEAD, P. 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research,* 47**,** n/a-n/a.

WOODS, R. A. 2009. Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. *Advances in Water Resources,* 32**,** 1465-1481.

YAMAZAKI, D., TRIGG, M. A. & IKESHIMA, D. 2015. Development of a global~ 90 m water body map using multi-temporal Landsat images. *Remote Sensing of Environment,* 171**,** 337-351.

YE, S., LI, H.-Y., LEUNG, L. R., GUO, J., RAN, Q., DEMISSIE, Y. & SIVAPALAN, M. 2017. Understanding Flood Seasonality and Its Temporal Shifts within the Contiguous United States. *Journal of Hydrometeorology,* 18**,** 1997-2009.

YUE, S., PILON, P. & PHINNEY, B. O. B. 2003. Canadian streamflow trend detection: impacts of serial and cross-correlation. *Hydrological Sciences Journal,* 48**,** 51-63.

YUE, S. & WANG, C. 2002. The influence of serial correlation on the Mann–Whitney test for detecting a shift in median. *Advances in Water Resources,* 25**,** 325-333.

ZAHERPOUR, J., GOSLING, S. N., MOUNT, N., SCHMIED, H. M., VELDKAMP, T. I. E., DANKERS, R., EISNER, S., GERTEN, D., GUDMUNDSSON, L. & HADDELAND, I. 2018. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environmental Research Letters*.

ZHAN, C., NIU, C., SONG, X. & XU, C. 2012. The impacts of climate variability and human activities on streamflow in Bai River basin, northern China. *Hydrology Research,* 44**,** 875-885.

ZHANG, A., ZHENG, C., WANG, S. & YAO, Y. 2015. Analysis of streamflow variations in the Heihe River Basin, northwest China: Trends, abrupt

changes, driving factors and ecological influences. *Journal of Hydrology: Regional Studies,* 3**,** 106-124.

ZHANG, X., HARVEY, K. D., HOGG, W. D. & YUZYK, T. R. 2001. Trends in Canadian streamflow. *Water Resources Research,* 37**,** 987-998.

ZHAO, F., VELDKAMP, T. I., FRIELER, K., SCHEWE, J., OSTBERG, S., WILLNER, S., SCHAUBERGER, B., GOSLING, S. N., SCHMIED, H. M. & PORTMANN, F. T. 2017. The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters,* 12**,** 075003.

ZHAO, G., LI, E., MU, X., WEN, Z., RAYBURG, S. & TIAN, P. 2015. Changing trends and regime shift of streamflow in the Yellow River basin. *Stochastic Environmental Research and Risk Assessment,* 29**,** 1331-1343.

ZISKIN, D., BAUGH, K., HSU, F.-C. & ELVIDGE, C. D. 2010. Methods used for the 2006 radiance lights. *Proceedings of the Asia-Pacific Advanced Network,* 30**,** 131-142.

ZVOLEFF, A. 2015. gfcanalysis: Tools for Working with Hansen et al. Global Forest Change Dataset. Available online at https://cran.r-project.org/web/packages/gfcanalysis.